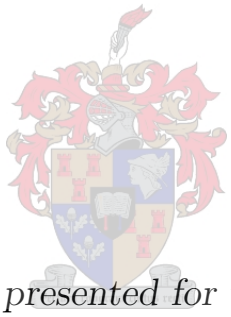
Characterisation of the South African Extreme Wind Environment Relevant to Standardisation

by

Frederik Pierre Bakker



Dissertation presented for the degree of

Doctor of Philosophy in Civil Engineering

in the Faculty of Engineering at Stellenbosch University

Supervisors: Prof. C. Viljoen

Dr. N. De Koker

March 2021

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

This dissertation includes three original papers intended for publication in peer-reviewed journals: one is in an advanced stage of review, one has been submitted for review, and one is ready for submission. The development and writing of the papers were the principal responsibility of myself and declarations are included in the dissertation indicating the nature and extent of the contributions of co-authors.

March 2021
Date: .....

Abstract

To ensure adequate structural resistance, engineers estimate the largest loads that wind can reasonably be expected to induce in a structure. In South Africa, operational design practices are described in a set of standards, which include a wind loading standard (SANS 10160-3) that provides guidance to engineers in estimating wind loads. The estimation procedure described in the standard uses a basic wind speed, provided in a characteristic map, that accounts for what the climate could produce at a particular location. This map is based on inference from available wind speed observations using stochastic models.

In an effort to improve the background information on which the standard is based and in response to greater volumes of wind speed data, this dissertation presents a set of investigations which incorporate a greater quantity of relevant data, strive to improve the stochastic model, and make specific recommendations that could be included in the next revision of the South African wind load standard.

A background study that summarises some of the most salient theory is also presented. This includes an overview of the South African extreme wind climate, extreme value theory, wind load standardisation, and reliability-based design.

The quantity of available data collected by the South African Weather Service has increased substantially since the current standard was derived. This data had to be checked for errors and classified by climatic mechanism. Following an attempt to correct for the dynamic response of outdated instruments, the inclusion of historical data dating as far back as 1948 was also investigated. To be relevant for use in the standard, all the data had to be measured according to World Meteorological Organisation standards. This was typically not the case, and so surface roughness correction was applied using a Geographical Information System technique.

The expanded dataset includes data from an extensive network of 132 stations throughout South Africa, but the use of only annual maxima resulted in insufficient sample sizes at individual sites, potentially introducing substantial model-variance. In response, improved estimates by incorporating data from multiple sites within the region of study were determined in this thesis. This included regional estimation of the shape of the probability distribution, based on preconditioning the data by an exponent; and a separate investigation on optimal estimation of variability. For estimation of the variability, it was shown that there was a balance between site and regional estimation, which was used to develop an optimal estimator based on the bias variance trade-off.

The expanded dataset and the insights into regional estimation were then applied to estimate design wind speeds that should satisfy the target reliability specified in the South African standard. These estimates were incorporated into a new characteristic wind speed map using a Monte Carlo simulation-based hypothesis test.

Oorsig

Om 'n struktuur te onwerp met voldoende weerstand, voorspel ingenieurs die kritiese windlading wat op die struktuur kan inwerk. In Suid-Afrika word die ontwerp prosedures beskryf in 'n stel standaarde wat 'n windlading standaard insluit (SANS 10160-3) waarin die voorspelling van windlading beskryf word. Hierdie voorspellings prosedure is gebaseer op 'n basiese wind spoed, wat op 'n windkaart beskyf is, wat voorsiening maak vir die klimaat by spesifieke liggings. Hierdie windkaart is gebaseer op afleidings van beskikbare wind spoed waarnemings en stogastiese modelle.

In 'n poging om die Suid-Afrikaanse windlading standaard te hersien, voorsien hierdie verhandeling 'n stel studies wat meer data gebruik om die stogastiese model te verbeter, asook nodige aanbevelings wat in die volgende hersiening van die Suid-Afrikaanse windlading standaard inkorporeer kan word.

'n Agtergrond studie wat 'n opsomming van die belangrikste teorieë bevat word ook aangebied. Dit bevat 'n oorsig van die Suid-Afrikaanse uiterste/sterk windklimaat, statistiese benaderings om uiterstes te beraam, standaardisering van windladings asook betroubaarheids-gebasseerde ontwerpe.

Die hoeveelheid data wat deur die Suid-Afrikaanse weerdiens versamel is, het aansienlik toegeneem sedert vorige studies wat gebruik was om die huidige standaard te ontwikkel. Hierdie data moes deur 'n kwaliteitsbeheer proses gaan en volgens verskillende tipe uiterste/sterk windklimate geklassifiseer word. Na 'n poging om die dinamiese reaksie van ouer instrumente in ag te neem, is die insluiting van windspoed meetings van so ver terug as 1948 ondersoek. Om vir die standaard gebruik te kan word, moet al die data volgens die Wêreld Meteorologiese Organisasie se standaarde gemeet word. Omdat dit nie gewoonlik die geval was nie, moes 'n korreksie van die grondwrywing toegepas word met behulp van Geografiese-Inligting-Stelsel tegnieke.

Die finale datastel bestaan uit data vanaf 'n netwerk van 132 weerstasies regoor Suid-Afrika. Die gebruik van jaarlikse maksima beteken dat individuele weerstasies probleme ervaar met klein steekproef grotes, wat onaanvaarbare model-variansie kan lewer. As gevolg van die bogenoemde, is data van verskeie weerstasies in die omgewing gebruik om die voorspellings te verbeter. Die toepaslike streeksverspreidings is ook bepaal, wat gebaseer is op die metode van maksimum waarskynlikheid. In terme van die beraming van veranderlikheid, dit getoon dat daar 'n balans tussen individuele en streeks beraming bestaan, wat gebruik is om die optimale beramer te ontwikkel.

Die verlengde datastel en insigte oor streeks voorspelling is gebruik om die ontwerp wind snelhede te bepaal wat die betroubaarheid van die Suid-Afrikaanse standaarde sal onderhou. Hierdie voorspellings is dan opgeneem in 'n nuwe windspoedkaart met behulp van 'n hipotesetoets gebaseer op Monte Carlo simulasies.

Acknowledgements

Firstly, I am grateful for the funding from the Stellenbosch University Institute for Structural Engineering, without which this work would not have been possible.

Secondly, wind speed data used in this study was provided by the South African Weather Service, where Dr Andries Kruger kindly assisted in the extraction of the data and gave much needed guidance during the verification and classification of gusts.

Finally, I would like to thank: Prof. Celeste Viljoen, who by giving me the opportunity to do this work, has set my life on a more interesting trajectory; Dr. Nico De Koker, for coaxing coherent thoughts out of my often muddled mind; my Father, for convincing me that engineers may in fact need statistics; and my Mother and Eugene, for the support and Sunday evening braais.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Engineers aim to design infrastructure with minimum wasted effort and expense. To this end, they tend to use deterministic models that approximate the behaviour of the real world.

To entirely model the behaviour of a complex real world system is rarely feasible, therefore engineers use simplified models which adequately predict outcomes to ensure safety and efficiency. The use of simplified models raises questions about the extent to which a model aligns with reality, thus by refining a model, a better reflection of reality could be achieved, resulting in designs with improved safety and efficiency.

A notoriously complicated system is the extreme wind climate, which produces storms with the potential to inflict substantial structural damage. Models that aim to predict these wind loads are described in loading standards that strive to allow engineers to efficiently allocate materials and achieve sufficient structural resistance. The South African wind loading standard (SANS 10160-3) is used to account for South African conditions.

1.1 Wind Loading Standards

A wind loading standard aims to predict the loads that a structure will experience throughout its operational lifetime. The most common model used is the Davenport

(1982) wind load chain (Holmes, 2018)

$$q_d = \frac{1}{2}\rho\gamma_w c_k v_k^2, \tag{1.1}$$

where wind load q_d is calculated from various components: the air density ρ, modification factors which account for terrain, aerodynamics and other effects c_k, a characteristic basic wind speed v_k, and a partial safety factor that accounts for uncertainty γ_w.

If these codified predictions are underestimated, structures may have insufficient resistance, potentially resulting in significant property damage, injury, and death. Conversely, if the predictions are overestimated, more resistance than required for a level of safety deemed acceptable by society, will be assigned, potentially resulting in substantial waste of resources.

Models in wind loading standards account for how key components such as the structural shape, local environment, and wind climate impact the wind loads on a structure. The availability of new wind speed data creates an opportunity to review and possibly refine parts of the existing model.

1.2 Reviewing the Representation of the Wind Climate in the South African Standard

Predictions of wind speeds with the potential to cause structural damage are represented in loading standards by *basic wind speeds*. These reference values are used in the model to predict wind loads, which are multiplied by a partial safety factor that aims to ensure a satisfactory level of safety. Both the reference values and the partial safety factor are determined using stochastic models based on the available data.

The infrequency of extreme wind speeds relevant to structural design, means recordings of these events are scarce. The availability of more data would create an opportunity to review existing stochastic models and therefore reconsider the reference wind speeds and/or the partial safety factor in the South African wind loading standard (SANS 10160-3).

This dissertation presents a series of investigations that aim to review the prediction of wind loads in (SANS 10160-3:2019). This could be achieved by:

- Increasing the quantity of relevant data (Chapter 3).
- The development of an appropriate spatiotemporal stochastic model incorporating the new data, (Chapter 4 and Chapter 5).
- A review of the existing standard, in light of the new data and stochastic model (Chapter 6).

These investigations may also contribute to wind engineering by:

- Streamlining the preparation and extraction of normalised wind speed data (Chapter 3).
- Improving the probabilistic representation of extreme events over a large region (Chapter 4 and Chapter 5).
- Developing a framework to simplify the incorporation of new information into a standard (Chapter 6).

1.3 Scope

This study is limited to the characterisation of the basic wind speed in Davenport (1982)'s wind load chain, with a specific focus on insights provided by increased volumes of data across a large spatial domain. Thus, a review of the probabilistic representation of the wind loading standard (Botha *et al.*, 2018*b*) is only considered in terms of basic wind speeds. Additional modification factors not directly informed by wind speed data, such as the factors that account for height and structural shape, are not considered.

Extreme events with the potential to cause structural damage are accounted for in the standard by ultimate limit state conditions (Retief and Dunaiski, 2009). Predictions of these rare events are more dependant on the stochastic model than predictions used for ensuring adequate function of the structure (serviceability limit state). Therefore, in

this project, only ultimate limit state conditions are explicitly considered, although the methods that are developed could still be applied for the serviceability limit state.

1.4 Structure of the Dissertation

The individual investigations in this dissertation are organised into chapters. These build upon the results of the previous investigations towards practical recommendations that could be adopted by SANS 10160-3.

Chapter 2 presents an introduction to wind loading standards and details on the current South African provisions. It describes how wind speed observations are incorporated into loading standards, emphasising the key role of extreme value theory.

Chapter 3 describes the collection and quality control of data measured by the South African Weather Service. This includes details on the identification of erroneous measurements, classification of gusts by dominant extreme wind mechanism, normalisation to reference wind speeds, and correction of instrumentation response.

Chapter 4 investigates appropriate probability distributions. The Gumbel distribution is suggested, but its relative inflexibility means that estimates of design values could be significantly biased. The use of a maximum likelihood method to justify models based on wind speed raised by an exponent is investigated.

Chapter 5 investigates the optimal use of the available data. Spatial variation of the wind climate means it is unclear which data are relevant at a particular location. This problem is framed in terms of the bias variance trade-off and a balance between site and regional estimation is pursued.

Chapter 6 combines the preceding investigations: the estimation procedure from chapter five, is applied to the distribution from chapter four, using the data assembled in chapter three, to give practical recommendations for the South African standard.

Chapter 7 concludes the research by summarising the work that has been conducted, limitations of the work, and suggesting areas for future study.

Chapter 2

Background

Summary

The overarching theme of this research is a review of how the wind climate is represented in the South African wind loading standard. This background chapter aims to provide context by reviewing South African strong wind climatology and its representation in the South African wind loading standard, emphasising the key role of extreme value theory.

Notation

q	Pressure
v	Wind speed
y	Height
ρ	Air density
q_d	Design wind pressure
v_k	Characteristic wind speed
γ_w	Wind speed partial safety factor
$F_v(v)$	Cumulative density function (CDF) of wind speed
F_{max}	CDF of the extreme value
n'	Number of events in a block
n	Number of extreme events (block maxima)
u_x	GEV location parameter
b_x	GEV scale parameter
k_x	GEV shape parameter
p	Exceedance probability
x	Generic random variable
$h(x)$	Exponential type function of x.
μ	Mean
σ^2	Variance
$\hat{u}_x$	Estimate of u
ll_{max}	Maximum log-likelihood
h	Number of parameters
ε_{tot}	Root mean square error
ε_{bias}	Model-bias error
ε^2_{std}	Model-variance error
ε^2_{irr}	Irreducible error
t	Return period
x_t	The t return period of x
$\hat{x}_t$	Estimate of t return period of x
e	Loading
r	Resistance
r_k	Characteristic value of resistance
Γ	Global safety factor
p_t	Target exceedance probability
l	Performance function
p_f	Probability of failure
f_e	Probability density function of e
g	Permanent loading
r_d	Design value of resistance
c	Wind load modification factor
γ_r	Partial safety factor of resistance
z	Standard normal variable
Φ	Standard normal CDF
$\mathbf{z}_d$	Vector of standard normal variables
β	Reliability index
α	Sensitivity factor
η	Ratio of wind to total load
γ	Euler's constant (0.57721...)
$g(x_i)$	Standardised Gumbel variate

2.1 Introduction

The destructive capacity of the wind is demonstrated by the damage that some storms can cause. A summary of wind related damage in South Africa is presented by Goliger and Retief (2007). Following analysis of storm event data collected by the National Oceanic and Atmospheric Administration (National Oceanic and Atmospheric Administration, 2020), it was found that between 1995 and 2020 around 3 500 deaths, 35 000 injuries and over 60 Billion US dollars of property damage could be attributed to strong winds throughout the United States.

Most of these costs were not associated with structural failure, and were the result of trees being uprooted, wind blown debris, and power line damage. Structures normally have adequate capacity because engineers specifically consider the wind loads that the structure could experience. Engineers are aided in this task by wind loading standards which provide guidance on the typical loads that wind could exert on a structure. To design for a wind load, an estimate of its magnitude at a particular point on a structure is required.

Consider Bernoulli's principle for laminar airflow

$$q_0 + \frac{1}{2}\rho v_0^2 + \rho g y_0 = q_1 + \frac{1}{2}\rho v_1^2 + \rho g y_1, \tag{2.1}$$

where q is the pressure, v is the velocity, y is the height, and ρ is the density of the air. Subscript 0 denotes the time before air is affected by the structure and subscript 1 denotes the time when air reaches the structure. When horizontally moving air ($y_0 = y_1$) is blocked by the structure and stagnates $v_1 = 0$, the resulting change in pressure can be represented in terms of the wind speed v_0 as

$$q_1 - q_0 = \frac{1}{2}\rho v_0^2, \tag{2.2}$$

which is termed the dynamic pressure. The load experienced by the structure can then be calculated by modifying the dynamic pressure to account for aerodynamics.

This has been adapted for the calculation of the design wind load q_d in standards (Holmes, 2018), which have almost universally adopted the Davenport (1982) wind load chain. The q_d can be represented as

$$q_d = \frac{1}{2}\rho\gamma_w c_k v_k^2, \tag{2.3}$$

where ρ is inferred from altitude and considered deterministic, $c_k = c_{k1}c_{k2}...c_{km}$ represents the combination of m modification factors which account for terrain, aerodynamics and other effects, v_k is the characteristic basic wind speed, and γ_w is the partial safety factor that accounts for uncertainty. The focus of this research is the wind climate, which is explicitly represented in this formulation by v_k, and implicitly represented by γ_w through the uncertainties in v_k that it accounts for.

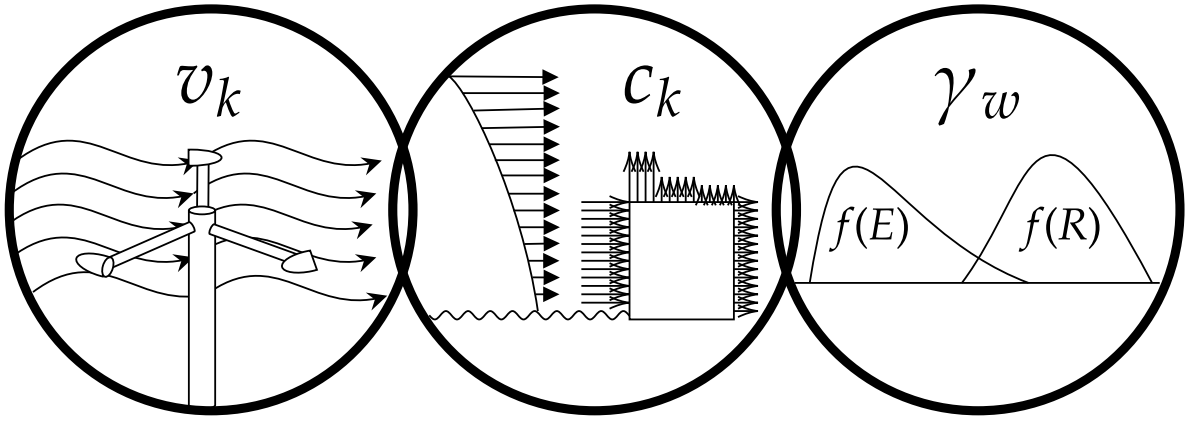


Figure 2.1: Wind load components.

2.2 South African Extreme Wind Climate

Wind speeds, and their associated pressures, are caused by numerous climatological systems. Those likely to damage property and human life are relevant to structural

design. Holmes (2018) identifies extratropical cyclones, tropical cyclones, thunderstorm downbursts and tornadoes as the main wind mechanisms that cause significant damage worldwide. All of these strong wind producing mechanisms have been observed at some point in South Africa (Kruger, 2011; Goliger and Retief, 2007).

2.2.1 South African Strong Wind Mechanisms

2.2.1.1 Synoptic

In mid-latitudes (40°-60°) a low pressure band around the Earth leads to the formation of synoptic scale extratropical cyclones, potentially thousands of kilometres in extent. These systems move with the prevailing westerly winds and generate cold fronts, with wind speeds of substantial strength, that can extend over a thousand kilometres and take days to pass.

In South Africa a low pressure band, with associated extratropical cyclones, is situated off the southern coast. During summer the Atlantic high pressure system lies over South Africa and blocks approaching cold fronts from these extratropical cyclones. In winter, this high pressure system shifts northwards and severe winds from cold fronts regularly reach southern parts of the country (Figure 2.2).

Kruger *et al.* (2010) found that extratropical cyclones along with four other synoptic scale mechanisms, tend to dominate the extreme wind climate over the south-western and coastal regions of South Africa.

2.2.1.2 Tropical Cyclones

Tropical cyclones are generated by uplift of air over warm oceans at tropical latitudes. These then normally move east to west with the prevailing winds and can cause significant damage when they make landfall, fortunately they tend to dissipate fairly quickly as they move inland.

Off the east coast of South Africa, the warm Indian Ocean frequently experiences tropical cyclones. The paths of these storms have been recorded by the Australian Bureau

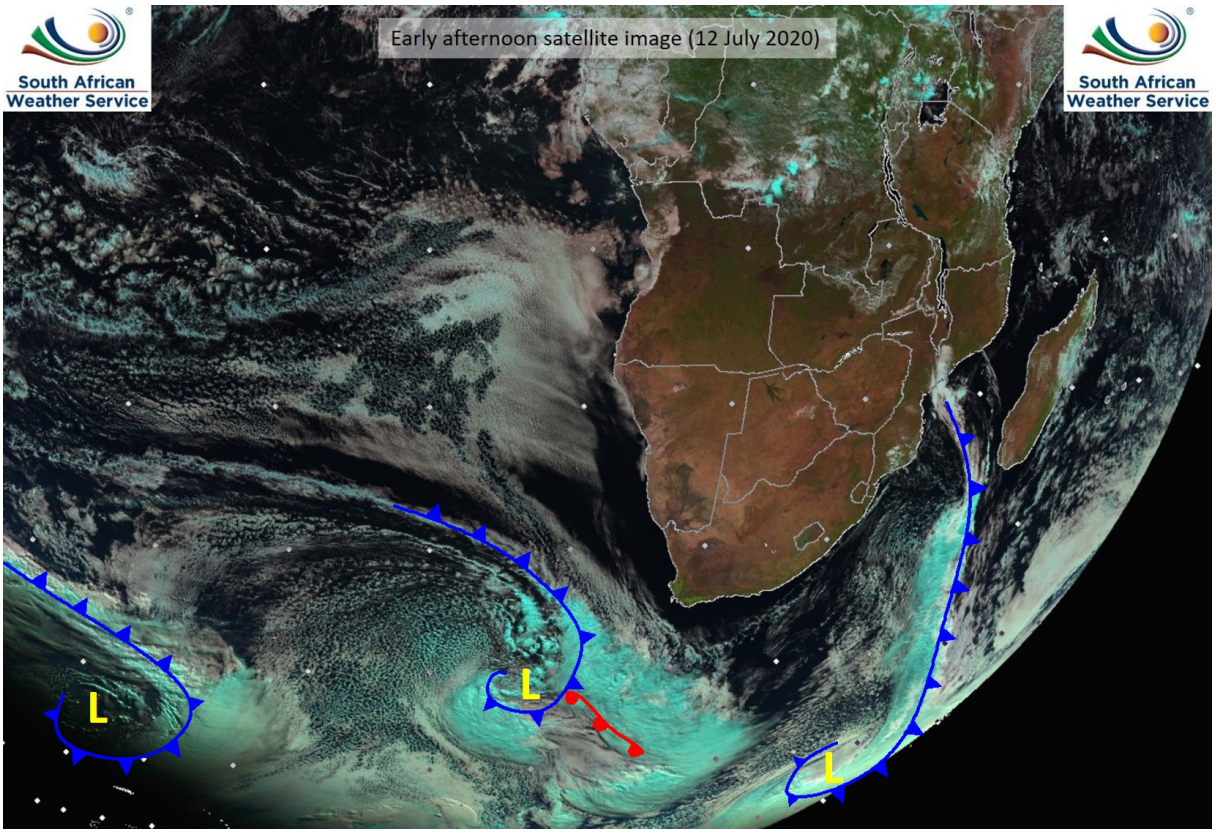


Figure 2.2: Extratropical cyclones near South Africa on 12 July 2020 with cold fronts indicated in blue (South African Weather Service, 2020).

of Meteorology, who keep a database of tropical cyclones in the southern hemisphere. This shows that since 1970 four storms have affected eastern parts of South Africa, shown in Figure 2.3. The frequency and intensity of severe tropical cyclones may increase due to climate change (Knutson *et al.*, 2010), this could explain why two of the tropical cyclones that managed to reach South Africa occurred in the last decade.

Kruger (2011) acknowledged the possibility of tropical cyclones in South Africa, although the risk of structural damage was deemed insignificant because when these storms have reached South Africa they have been in the process of dissipating.

2.2.1.3 Thunderstorms

Thunderstorm downbursts are localised convective systems generally only a few kilometres in extent and are caused by the uplift of warm moist air over land. This air cools rapidly as it rises and condenses until it becomes too heavy and begins to move downwards. When the air reaches the earth's surface it spreads out and forms a gust front with high

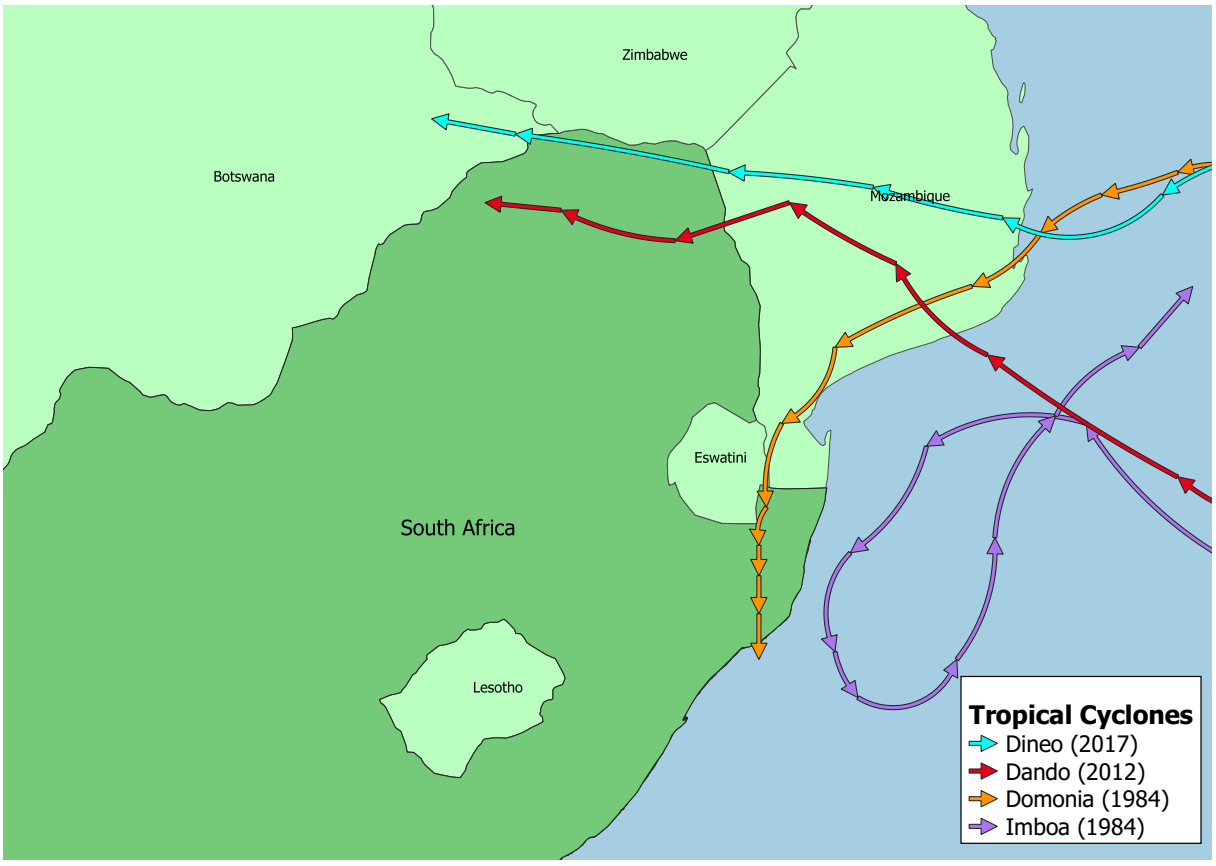


Figure 2.3: Tropical cyclones in South Africa (1970 - 2019) (Australian Government, Bureau of Meteorology, 2020).

wind speeds that last for several minutes and can cause substantial damage (Goliger and Retief, 2007).

During summer, moist air from the Indian Ocean high pressure system is trapped over the South African interior. This air then rises due to convection or topography which leads to strong thunderstorm downbursts over much of the South African interior (Kruger *et al.*, 2010).

2.2.1.4 Tornadoes

Another type of convective system which could cause significant damage are tornadoes. These are characterised by a funnel vortex and are small, typically only a few hundred metres, although they can travel for long distances, up to 50 kilometres (Holmes, 2018).

A dataset of over 330 South African tornadoes has been compiled by Miller (2020). This dataset built on a previous study by Milford *et al.* (1994) using various sources

including news reports, Youtube videos, and social media. The strength of each tornado has been estimated using the Fujita-Pearson classification system (Fujita, 1971). In total; 55 category F2, 25 category F3, and 3 category F4 tornadoes were identified. A map of these tornadoes with approximate wind speeds is shown in Figure 2.4. The wind speeds from these tornadoes are significantly higher than specified by SANS 10160-3:2019 and thus have the potential to cause significant damage, such as the F4 tornado in 2011 that levelled houses in Duduza township in Johannesburg.

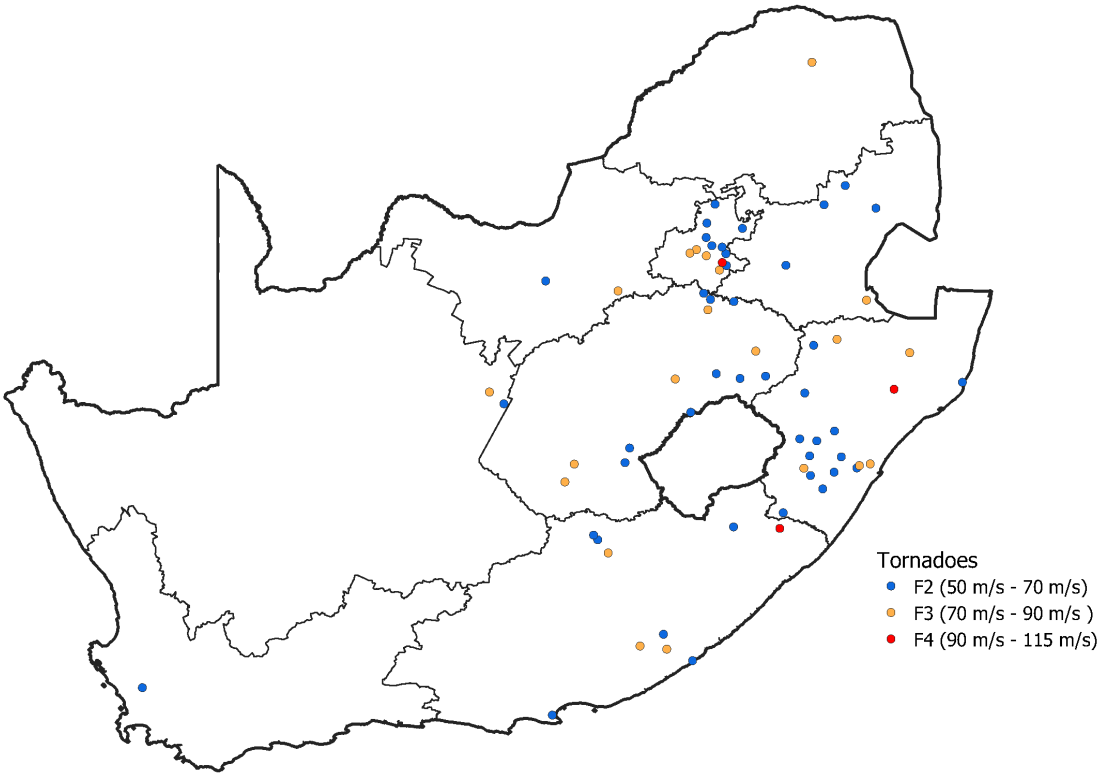


Figure 2.4: Recorded tornadoes in South Africa (1905-2019).

2.2.2 Strong Wind Mechanisms Relevant to South African Structural Design

Tropical cyclones have rarely entered South Africa and there is little evidence of wind induced damage when they have. The probability of a tornado affecting a structure covered by SANS 10160-3:2019 is low, in part due to the small spatial extent of these

events (Milford *et al.*, 1994). Designing a structure to withstand a tornado would add significant cost.

It is argued that tropical cyclones and tornadoes can be neglected in this study on the grounds of probabilistic optimisation (Holickỳ, 2009). Tropical cyclones have a low probability of occurrence and low damage potential, and tornadoes have a low probably of occurrence and high cost of safety.

Of the four main damage-inducing wind mechanisms, Kruger (2011) only considered synoptic and thunderstorm winds for the derivation of basic wind speeds in SANS 10160-3:2019. These winds occur frequently and are the most probable source of extreme winds throughout South Africa (Goliger and Retief, 2007). Kruger (2011) found that thunderstorms dominated the north eastern and synoptic winds dominated the south-western regions of South Africa, as shown in Figure 2.5.

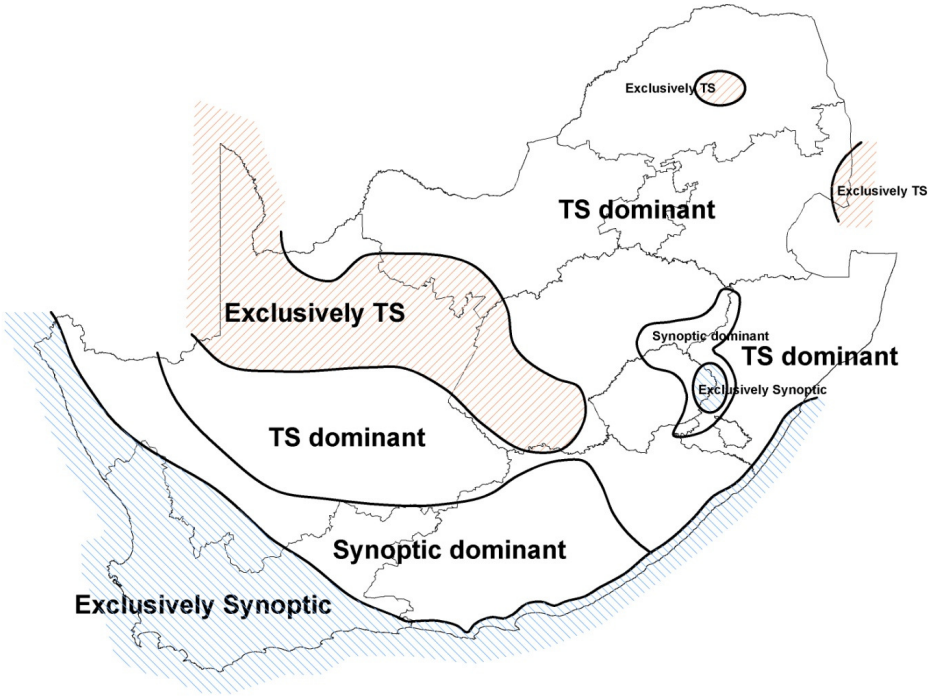


Figure 2.5: Dominant extreme wind speed mechanisms in South Africa (Kruger, 2011).

2.3 Extreme Value Theory

National weather organisations continuously collect and record wind speeds, this results in a large quantity of data which could be used to draw conclusions about the strongest wind speed likely to occur over the life of nearby structures. Generalising the wind speed probability distribution allows meaningful conclusions to be drawn about what constitutes an adequate wind speed for structural design and is a key link between the observed wind record and the values used by engineers in standards.

The basic wind speeds v_k and the partial factor γ_w are both dependant on the statistical distribution used to represent the wind speeds v. The distribution that is fit should be appropriate for the quantity of representative data that is available. A naive approach would be to fit a distribution to all the available data to estimate the cumulative density function of the wind speeds $F_v(v)$ and then assume that wind speeds are independent and identically distributed. The cumulative distribution of the maximum value $F_{max}(v)$ of n' observations is then

$$F_{max}(x) = F(x)^{n'}. \tag{2.4}$$

This is inappropriate because wind speed recordings are not serially independent or identically distributed.

Wind speeds observed in temporal proximity are likely to be autocorrelated which violates the assumption of independence. This cannot be reliably accounted for yet, because the degree and rate of correlation is not clearly defined (Harris, 2017; Torrielli *et al.*, 2016, 2017). This issue is further complicated when considering multiple mechanisms as the nature of the correlation will likely differ depending on the underlying mechanism.

This relates to the second flaw, which is that the underlying mechanism can differ substantially and may not be similarly distributed. Because the strong winds discussed in Section 2.2 are comparatively rare, any analysis will tend to be dominated by irrelevant information and fail to make good predictions, unless the relevant extreme winds are specifically considered.

Extreme value theory has been developed to deal with these issues.

2.3.1 Block Maxima and Asymptotic Distributions

A central principle of extreme value theory is the specific analysis of independent extremes separate from the parent population. These extremes can be extracted using a block maximum approach, where the observed record is divided into time periods, termed blocks, of equal size and only the maximum from each block is considered further. This procedure is demonstrated graphically in Figure 2.6. If the size of the block is large

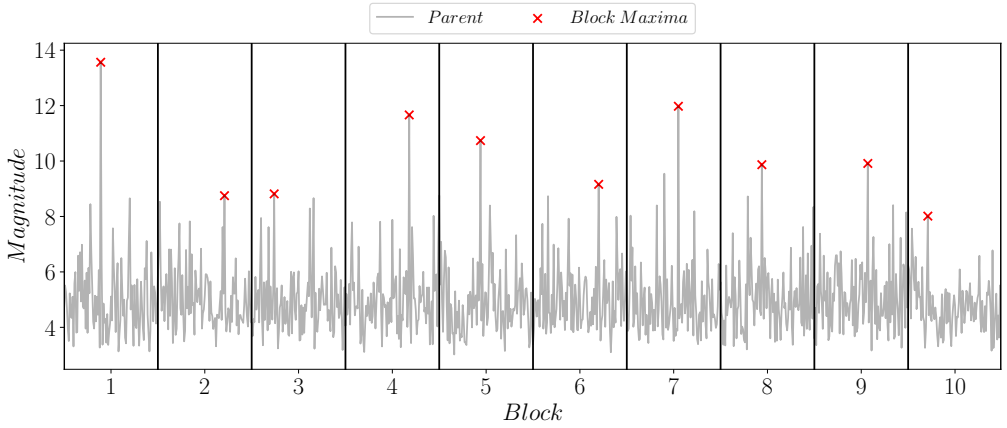


Figure 2.6: Block maximum approach.

enough, then taking the maximum yields only extreme values which should be independent given sufficient separation between maxima. For wind speeds, a block size of one year is convenient as seasonal fluctuations can be neglected. Characterising the distribution of these block maxima provides information about the probability of the largest value that could be expected, which for ultimate limit state loads is the value of greatest interest.

Fisher and Tippett (1928) showed that the distribution of block maxima F_{max} approach one of three limiting forms as the number of events per block $n' \to \infty$. These asymptotic forms were combined by Von Mises (1936) as the Generalised Extreme Value (GEV) distribution;

$$F_{max}(x) = \begin{cases} \exp\left[-\exp\left(-\frac{x-u_x}{b_x}\right)\right] & k_x = 0 \\ \exp\left[-\left(1-k_x\frac{x-u_x}{b_x}\right)^{\frac{1}{k_x}}\right] & k_x \neq 0. \end{cases} \tag{2.5}$$

The GEV distribution has a location, scale and shape parameter denoted by u_x, b_x and k_x, respectively. The k_x parameter corresponds to the asymptotic form with $k_x = 0$ indicating the Gumbel/Type 1 distribution, $k_x < 0$ the Fréchet/Type 2 distribution and $k_x > 0$ the Reverse Weibull/Type 3 distribution. The asymptotic form is dependant on the tail behaviour of the parent distribution. The block maxima from a parent with an exponential tail converge to a Gumbel distribution, maxima from a parent that decays with a polynomial tail will converge on a Fréchet distribution and a bounded distribution yields Reverse Weibull distributed extremes.

The asymptotic form arises as $n' \to \infty$ and therefore, because n' is finite, convergence to the asymptotic form is never entirely complete. This is of little practical concern provided n' is large enough that errors due to insufficient convergence are small (Cook and Harris, 2004). The use of only one value per block, means that many $(n' - 1)$ measurements are not utilised and errors due to lack of data is often a concern. This issue is especially common for wind data where records are rarely greater than 70 years. Thus limiting the analysis to the Gumbel distribution ($k_x = 0$), to avoid the high model-variance (2.3.5) associated with estimating k_x, is often appropriate (Kruger, 2011; Hong *et al.*, 2014).

Various techniques that allow more extreme values to be extracted from less parent data have been developed (Palutikof *et al.*, 1999). These include: extending the number of extracted extremes per block to the r-largest; the Method of Independent Storms (MIS), which involves specific consideration of the time series to identify storms and analysing the maximum value from each storm; or using all values which exceed a certain high threshold, termed Peak-over-threshold (POT). The problem with these approaches is that while they can offer more stability, they require more careful application than asymptotic theory. For these techniques more subjective decisions need to be made, for example, the r value in the r-largest analysis, storm identification in MIS, or the threshold value in POT.

2.3.2 Accounting for Multiple Mechanisms

A series of annual maximum wind speeds may not be identically distributed because the extreme wind environment could be a result of m different extreme wind mechanisms, as discussed in Section 2.2.

The v fractile v_{joint} that results from the joint distribution of m mechanisms for a given exceedance probability p is unknown. This can be found if the extreme wind speed record can be segregated by mechanism, i.e an extreme from each mechanism extracted per block, then the joint distribution considered, as recommended by (Gomes and Vickery, 1978).

The v_{joint} is equal to the fractile of each marginal distribution v_i

$$v_{joint} = v_1(p_1) = v_2(p_2) = \dots v_m(p_m). \tag{2.6}$$

Provided each mechanism is independent, p can be related to the exceedance probability of each independent mechanism p_i as

$$(1-p) = (1-p_1)(1-p_2)\dots(1-p_m). \tag{2.7}$$

These m unknown p_i values and the unknown v_{joint} constitute $(m+1)$ unknowns.

Equations 2.6 and 2.7 yield a system of $(m+1)$ equations. Therefore, provided each v_i can be related to p_i, the $(m+1)$ equations can be used to solve the $(m+1)$ unknowns. This solution will provide the v_{joint} for a given p, as is typically required.

2.3.3 Gumbel Distribution

The Gumbel (1958) distribution ($k_x = 0$) has been widely applied in the analysis of extreme wind speeds (Botha *et al.*, 2018*a*; Kruger, 2011; Hong *et al.*, 2014; Holmes, 2018). It has the significant advantage over the other asymptotic forms that a shape parameter does not need to be estimated. This decreases the model-variance (defined in Section 2.3.5), which can be large when data is limited by only using annual maxima (block size

of one year). Given its pre-eminence, a derivation of the Gumbel distribution by Cramér (1946), given by Harris (1996) is presented.

A random variable x is of the exponential type if its cumulative probability distribution can be represented as

$$F(x) = 1 - \exp\left[-h(x)\right], \tag{2.8}$$

where $h(x)$ is a function of x that is positive and monotonically increases faster than $\ln(x)$, ie. rate of increase $> 1/x$. This includes many widely used distributions such as the Normal and Forwards Weibull distribution.

From Eq. 2.4 the distribution of the maximum of n' events is

$$F_{max}(x) = \left(1 - \exp\left[-h(x)\right]\right)^{n'}. \tag{2.9}$$

The reference largest value u_x is

$$1 - F(u_x) = \frac{1}{n} = \exp\left[-h(u_x)\right]. \tag{2.10}$$

Thus

$$F_{max}(x) = 1 - \left(\frac{\exp\left[h(u_x) - h(x)\right]}{n'}\right)^{n'}, \tag{2.11}$$

which is still exact, but n' is typically unknown. If n' is large, then using the Cauchy approximation $F_{max}(x)$ can be found by taking $\lim_{n'\to\infty} F_{max}(x)$ which yields

$$F_{max}(x) = \exp\left[-\exp\left(h(u_x) - h(x)\right)\right]. \tag{2.12}$$

Taking logarithms twice

$$-\ln\left[-\ln\left(F_{max}(x)\right)\right] = h(x) - h(u_x) \tag{2.13}$$

and expanding $h(x) - h(u_x)$ about u_x using a Taylor series gives

$$-\ln\left[-\ln\left(F_{max}(x)\right)\right] = (x-u_x)\frac{h'(u_x)}{1!} + (x-u_x)^2\frac{h''(u_x)}{2!} + \dots . \tag{2.14}$$

Truncating the series at the first term and setting $b_x = 1/h'(u)$ gives

$$-\ln\left[-\ln\left(F_{max}(x)\right)\right] = \frac{x-u_x}{b_x}, \tag{2.15}$$

which is equivalent to the Gumbel distribution. This derivation explicitly reveals two potential sources of error: one from the Cauchy approximation and the other from neglecting higher order terms in the Taylor series.

The error due to the Cauchy approximation is small provided the n' is large. While wind speeds are not serially independent, if the time between measurements is sufficiently large then wind measurements can be considered independent. If a block size of one year is used, then the number of effectively independent wind speeds measurements per year is n'.

The number of effectively independent wind speed events per year was initially estimated to be 150 (Cook and Harris, 2008), but could be as large as 17 000 (Torrielli *et al.*, 2013). This remains a topic of debate (Torrielli *et al.*, 2016; Harris, 2017; Torrielli *et al.*, 2017; Cook, 2018), but is generally accepted to be sufficiently large that the Cauchy approximation does not introduce significant error (Cook and Harris, 2004).

The error due to neglecting the higher order terms in the Taylor series can be significant (Cook and Harris, 2004), especially as $(x - u_x)$ increases, i.e as one extrapolates from u_x to rarer events as is typically required in wind load standards.

When $h(x) = ax$, i.e the parent distribution is exponential, the Cauchy approximation error is small and the error in truncating the Taylor series after the first term is zero, thus convergence to the Gumbel asymptote is rapid.

2.3.4 Parameter Estimation

To define a Gumbel distribution with a probability density function of

$$f_{max}(x) = \frac{1}{b_x} \exp\left[-\frac{x-u_x}{b_x} - \exp\left(-\frac{x-u_x}{b_x}\right)\right], \tag{2.16}$$

the parameters u_x and b_x need to estimated from a sample of available data $\{x_1, x_2, \ldots x_n\}$. Various approaches to parameter estimation have been proposed (Palutikof *et al.*, 1999; Hong *et al.*, 2013; Hong, 2013). Four well known methods are described here, and compared to one another in Section 2.3.5.

The fitting procedures are demonstrated on a simulated sample of size $n = 15$, drawn from a Gumbel distribution with $u_x = 20$ and $b_x = 3$. The sample and different fits are plotted on Figure 2.7 as described in Section 2.3.4.2. The sample represents only one realisation. To get a sense of all possible realisations 10 000 samples were simulated and the resulting distributions are plotted by rank in Figure 2.7.

2.3.4.1 Method of Moments

The mean μ_x can be estimated from a sample $\{x_1, x_2, \ldots x_n\}$ as

$$\hat{\mu}_x = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{2.17}$$

and the variance σ_x^2 estimated as

$$\hat{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_x)^2. \tag{2.18}$$

The μ_x and σ_x^2 of a Gumbel distribution can be found in terms of u_x and b_x as

$$\begin{aligned} \mu_x &= \int_{-\infty}^{\infty} x f_{max}(x) dx \\ \mu_x &= u_x + \gamma b_x, \end{aligned} \tag{2.19}$$

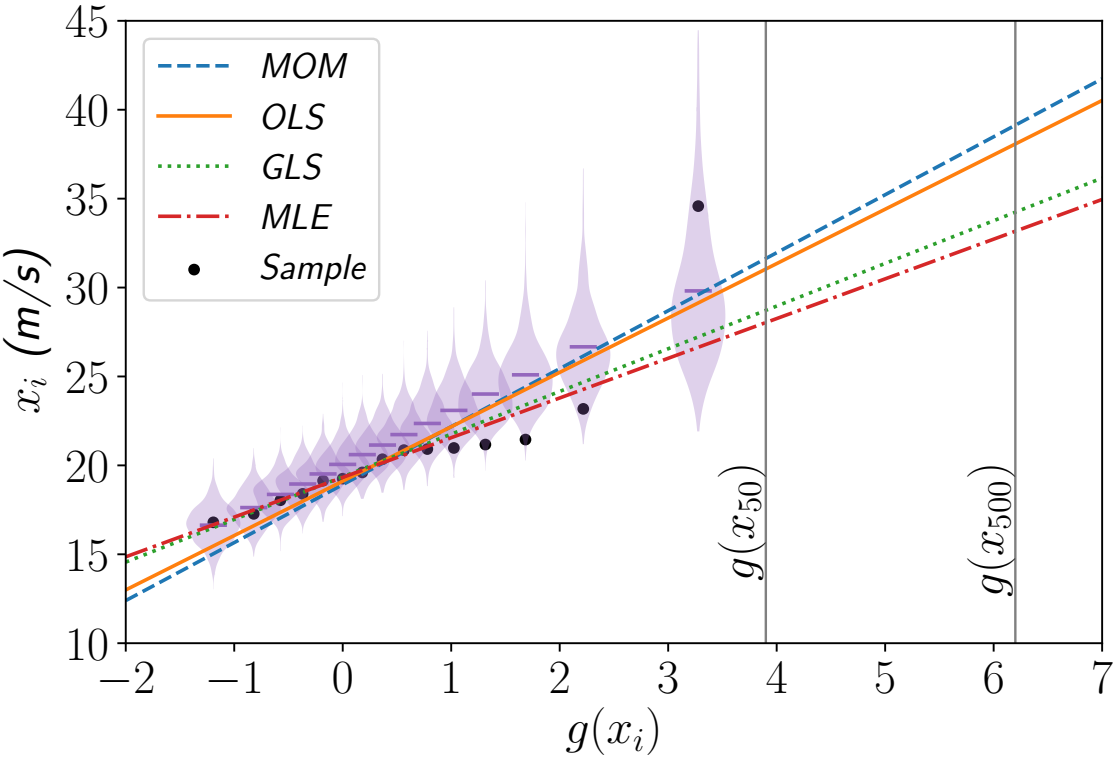


Figure 2.7: Parameter estimation techniques applied to a sample from a Gumbel distribution, with the distributions of order statistics shown ($n = 15$, $u_x = 20$, $b_x = 3$). Data is plotted using Gringorten (1963)'s plotting positions and the distributions of each rank were found using Monte Carlo simulation. The $g(x_i) = -\ln\left[-\ln\left(1-p\right)\right]$ is referred to as the standardised Gumbel variate and transforms the scale to extreme value probability paper, where p is the exceedance probability.

$$\begin{aligned} \sigma_x^2 &= \int_{-\infty}^{\infty} (x-\mu_x) f_{max}(x) dx \\ \sigma_x^2 &= \frac{\pi^2}{6} b_x^2, \end{aligned} \tag{2.20}$$

where $\gamma = 0.577\ldots$ is Euler's constant.

Combining Eqs. 2.20 and 2.18, b_x can be estimated as

$$\hat{b}_x = \frac{\sqrt{6}}{\pi} \hat{\sigma}_x. \tag{2.21}$$

which with Eqs. 2.19 and 2.17, also allows u_x to be estimated as

$$\hat{u}_x = \hat{\mu}_x - \gamma \hat{b}_x. \tag{2.22}$$

This is the method of moments (MOM), the resulting fit is shown on a Gumbel plot in Figure 2.7.

2.3.4.2 Graphical

A Gumbel distribution can be linearised by defining $g(x) = -\ln[-\ln(F_{max}(x))]$, then from Eq. 2.15, $g(u_x) = 0$ and the derivative $g'(x) = b_x$. Therefore, if x is Gumbel distributed, then $g(x)$ plotted against x_i would be a straight line with intercept u_x and slope b_x.

Given a series of n block maxima an empirical estimate of F_{max} can be obtained from the order statistics by ranking the sample from smallest to largest $\{x_1 < x_2 < \dots x_n\}$, then

$$\hat{F}_{max}(x_i) = \frac{i}{n+1}, \tag{2.23}$$

where i is the ranked position of x_i out of the n available extreme values. This empirical estimate is biased for the Gumbel distribution (Palutikof *et al.*, 1999), an alternative unbiased plotting position is given by Gringorten (1963) as

$$\hat{F}_{max}(x_i) \approx \frac{i - 0.44}{n + 0.12}. \tag{2.24}$$

If the estimate $\hat{g}(x_i) = -\ln\left[-\ln\left(\hat{F}_{max}(x_i)\right)\right]$ is plotted against x_i and a line fitted, then reasonable parameter estimates can be found as the intercept $\hat{u}_x$ and the slope $\hat{b}_x$ of the fitted line.

Gumbel (1958) initially proposed using an ordinary least squares fit (OLS) which finds the line that minimises $(\hat{g}(x_i) - x_i)^2$, demonstrated in Figure 2.7.

The issue with OLS, is that the variance is rank dependant, as shown in Figure 2.7, and there is correlation between ranks which is not accounted for. Assigning equal weight to each point thus increase the variability of the prediction and will result in greater average error (Section 2.3.5). This is demonstrated in Figure 2.7 where OLS is more affected by the largest value ($i = 15$) at $g(x_{15}) = 3.2$.

The variance and correlation can be accounted for by calculating the variance-covariance matrix for ranked data (Balakrishnan and Chan, 1992), and fitting using generalised

least squares (GLS) (Hong *et al.*, 2013). This is equivalent to the Best Linear Unbiased Estimation procedure developed by Lieblein (1976). The is demonstrated in the GLS fit in Figure 2.7 which is less affected by the largest value ($i = 15$) at $g(x_{15}) = 3.2$.

2.3.4.3 Maximum Likelihood Estimation

Another technique that does not suffer from the same drawbacks as OLS is maximum likelihood estimation (MLE). This aims to find the parameters for which $\{x_1, x_2, \ldots x_n\}$ has the highest likelihood of occurrence and is one of the most robust parameter estimation techniques, applicable to many distributions and situations.

The likelihood given the parameters $L(u_x, b_x)$ is found by evaluating the joint probability of the observed data sample

$$L(u_x, b_x) = \prod_{i=1}^{n} f_{max}(x_i|u_x, b_x). \tag{2.25}$$

Taking the log of the probability density is computationally convenient

$$\begin{aligned} ll(u_x, b_x) &= \sum_{i=1}^{n} ln[f(x_i, \theta)] \\ &= -n\ln(b) - \sum_{i=1}^{n} \frac{x - u_x}{b_x} - \sum_{i=1}^{n} \exp\left(-\frac{x - u_x}{b_x}\right), \end{aligned} \tag{2.26}$$

where $ll(u_x, b_x)$ is the log-likelihood function, whose maximum will correspond to the maximum of $L(u_x, b_x)$. Maximising this function gives the parameters with the highest likelihood of producing the observed data

$$\{\hat{u}_x, \hat{b}_x\} = \underset{\{u_x, b_x\}}{\text{argmax}}[ll(u_x, b_x)], \tag{2.27}$$

which are the maximum likelihood estimates of $\{u_x, b_x\}$ and can be found using a suitable numerical optimisation algorithm (Nelder and Mead, 1965; Snyman, 2005). The resulting fit is demonstrated on a Gumbel plot in Figure 2.7.

2.3.5 Model Selection

A model is used to describes the relationship between a set of input values and some output of interest. A good model aims to minimise the difference between the model output (prediction) and the actual values. This difference is the error of a model and can be represented by the root mean squared difference (RMSE) between a model prediction and the actual value.

The total RMSE ε_{tot} can be expressed in terms of three types of error (Friedman *et al.*, 2001)

$$\varepsilon_{tot} = \sqrt{\varepsilon_{bias}^2 + \varepsilon_{std}^2 + \varepsilon_{irr}^2}. \tag{2.28}$$

The ε_{bias} is the average difference between the model prediction and the actual value (model-bias), the ε_{std}^2 is the variance of the model predictions (model-variance) and ε_{irr}^2 represents irreducible errors outside the scope of available models. Since all models should be equally affected by ε_{irr}^2 it can often be neglected for the purposes of model selection.

2.3.5.1 Measure of Model Performance

Generally as a model becomes more complex (parameters increase), it is better able to fit the data and ε_{bias}^2 is decreased, but this increased flexibility often means the ε_{std}^2 increases.

Akaike (1973) quantified this trade-off using the Akaike Information Criterion (AIC)

$$AIC = -2ll_{max} + 2h, \tag{2.29}$$

which attempts to measure the relative information loss of competing candidate models. The better a model fits the data, the higher the maximum log-likelihood ll_{max}, and the lower the AIC. The greater the number of parameters h, the lower the certainty in a fitted model, and the higher the AIC. The AIC is used in model selection by selecting from a number of competing candidate models the one which yields the smallest AIC.

To measure the strength of the evidence for a particular candidate model m relative to the M models under consideration, the Akaike weight can be used (Burnham and

Anderson, 2002)

$$\text{Akaike Weight} = \frac{\exp[(AIC_{min} - AIC_m)/2]}{\sum \exp[(AIC_{min} - AIC_i)/2]}. \tag{2.30}$$

This can be interpreted as the probability, given the data, that model m is the best of the M models being considered. An adjustment to the AIC for small sample sizes is the AICc (Hurvich and Tsai, 1989)

$$AICc = -2ll_{max} + \frac{2hn}{n - h - 1}. \tag{2.31}$$

This increases the penalty for additional parameters when the sample size n is low. Burnham and Anderson (2002) recommend that a sample be considered small when $n/max(h) \lessapprox 40$.

The AIC has been used to compare different extreme value approaches applied to wind data (Hong *et al.*, 2014; Baravalle and Köhler, 2018), where AICc is typically appropriate if annual maxima are used.

2.3.5.2 Fitting Procedure Comparison

The performance of the fitting procedures discussed in Section 2.3.4 can be compared in terms of Root Mean Square Error (RMSE). Other than the obvious benefit of these comparisons to help decide which fitting procedure to use, this will also demonstrate some principles of model selection.

To test the fitting procedures, a Monte Carlo simulation procedure can be used to calculate the error:

- First m samples each of size n are drawn from a Gumbel distribution with known u_x and b_x.
- Then MOM, OLS, GLS and MLE is applied to each simulated sample to estimate the distribution parameters.
- The fitted distributions can be used to estimate errors of the t-year return period values x_t.

The model-bias is founds as

$$\varepsilon_{bias} = \frac{1}{m}\sum_{i=1}^{m}(\hat{x}_t - x_t), \tag{2.32}$$

and model-variance as

$$\varepsilon_{std}^2 = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{x}_t - \hat{\mu}_{x_t})^2, \tag{2.33}$$

where $x_t = -\ln[-\ln(1-1/t)]b_x + u_x$, and $\hat{x}_t$ is an estimate of x_t found using the $\hat{u}_x$ and $\hat{b}_x$ from one of the fitting procedures. There is no irreducible error $\varepsilon_{irr}^2 = 0$, and so ε_{tot} can be found with Eq. 2.28.

This was applied with $u_x = 20$, $b_x = 3$, $n = \{10, 20, 50\}$, and $t = \{50, 500\}$ which are representative of typical values in wind engineering (Holmes, 2018). The results are shown in Table 2.1, where $m = 10\ 000$ was used. Figure 2.8 is used to emphasise the reduction in RMSE as sample size increases for $t = 500$.

As expected, errors are greater for the higher return period and decrease as sample size increases. Generally MOM and MLE are moderately biased compared to the graphical procedures. Although, MLE makes up for this by being the most stable technique with the lowest ε_{std}^2. The overall predictive ability of the model, assessed through RMSE, show that GLS and MLE offer similar performance to one another and can be expected to perform better than MOM and OLS. A similar exercise, using more estimation procedures and a wider range of parameters, has been performed by Hong *et al.* (2013) and Hong (2013) that confirm these trends.

While GLS and MLE may be superior to less sophisticated techniques, model-variance is typically more of a concern (Holmes, 2018). In Figure 2.8 it is shown that the reduction in RMSE by increasing n from 10 to 20 is significantly larger than can be achieved by improving the fitting procedure.

		x_{50}			x_{500}		
		$n = 10$	$n = 20$	$n = 50$	$n = 10$	$n = 20$	$n = 50$
ε_{bias}	MOM	-0.51	-0.21	-0.10	-0.85	-0.47	-0.16
	OLS	-0.02	0.07	0.03	-0.10	-0.04	0.06
	GLS	-0.09	-0.02	-0.02	-0.25	-0.12	-0.03
	MLE	-0.81	-0.38	-0.16	-1.39	-0.68	-0.25
ε^2_{std}	MOM	14.30	8.08	3.20	33.54	18.10	7.48
	OLS	15.21	8.24	3.17	35.58	18.40	7.44
	GLS	11.99	6.03	2.24	27.14	13.08	5.05
	MLE	10.70	5.71	2.19	23.99	12.31	4.92
ε_{tot}	MOM	3.82	2.85	1.79	5.85	4.28	2.74
	OLS	3.90	2.87	1.78	5.97	4.29	2.73
	GLS	3.46	2.46	1.50	5.22	3.62	2.25
	MLE	3.37	2.42	1.49	5.09	3.57	2.23

Table 2.1: Model-bias, model-variance and RMSE of 50 and 500 year return period estimates using different parameter estimation techniques $\{MOM, OLS, GLS, MLE\}$ and sample sizes $n = \{10, 20, 50\}$.

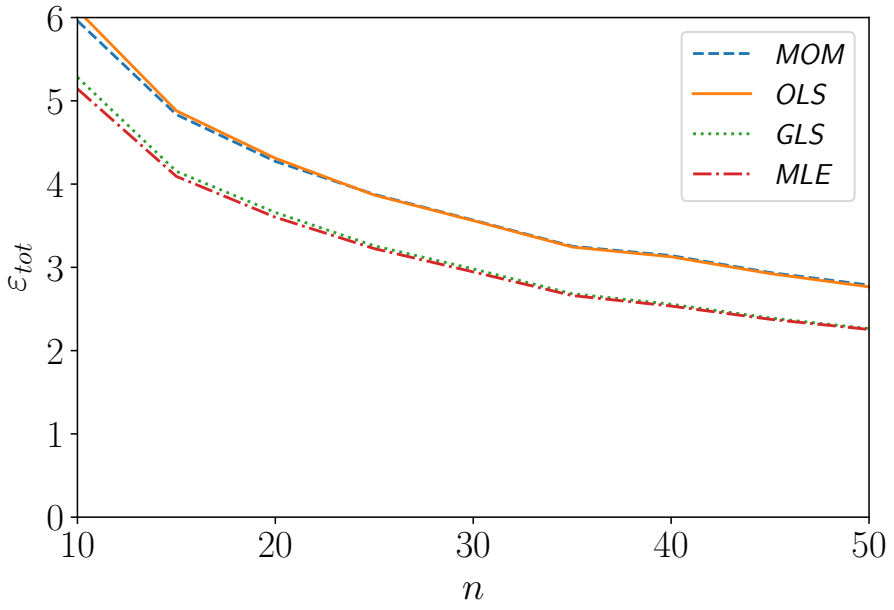


Figure 2.8: RMSE vs. n, for $\{MOM, OLS, GLS, MLE\}$ with $u_x = 20$, $b_x = 3$, and $t = 500$.

2.4 Basic Wind Speed

The differences in wind climate throughout the region covered by a standard is accounted for using the basic wind speed v_k. This should be based on relevant wind speed observations. In the South African standard (SANS 10160-3:2019), v_k is specified as the 50-year return period, 3-second gust wind speed, irrespective of wind direction and time of the year, measured at 10 metres above ground level, in open country terrain with low vegetation. This definition aligns with World Meteorological Organisation (WMO) standards so that wind speeds measured by meteorological organisations can serve to define v_k.

2.4.1 Measurement of Wind Speeds

The standardisation of weather stations is intended to ensure that measurements are not skewed by local conditions and are representative of the regional climate. In practice it is difficult to meet these standards for a sufficient distance and in all directions from a weather station. The WMO provides recommendations for correction of measurements made under non-standard conditions (World Meteorological Organization, 2008), which would need to be applied for many of the South African Weather Service (SAWS) stations (Kruger, 2011).

Modern national weather organisations automatically measure wind speeds using cup or vane anemometers (Brock *et al.*, 2001), such as the RM Young sensors currently used by the SAWS. These function by having the wind drive propellers that produce a voltage signal, with frequency directly proportional to the wind speed. Once calibrated, these devices should yield accurate measurements of the wind speed. Technical specifications for RM Young sensors suggest an accuracy 0.3 m/s for wind speeds up to 100 m/s (R.M. Young Company, 2005).

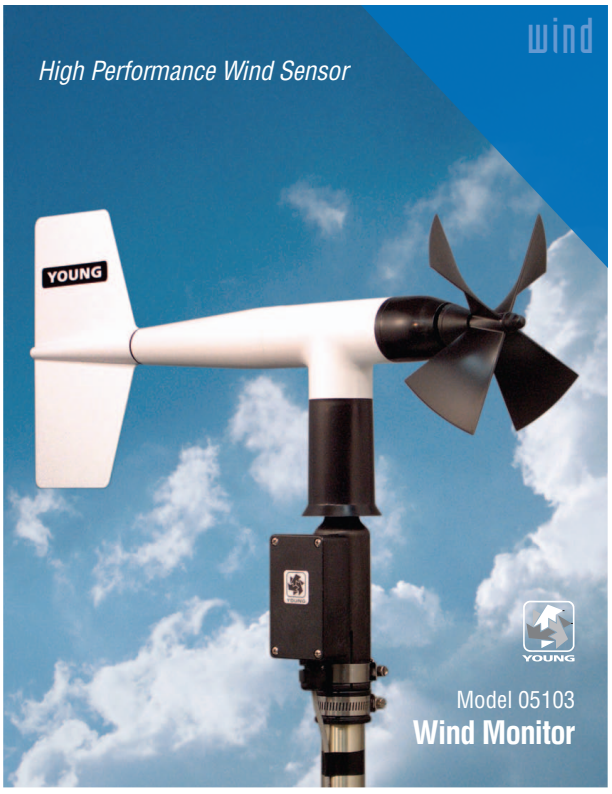


Figure 2.9: RM Young Sensor (R.M. Young Company, 2005)

2.4.2 Averaging Period

To decrease the impact of random wind fluctuations and the response of a specific instrument, recordings are smoothed over an averaging period. The WMO recommends that wind speeds should be sampled every 0.25 seconds and averaged for every 3 second period. This is justified because gusts that persist for at least 3 seconds could be capable of engulfing typical structures (World Meteorological Organization, 2008). Holmes and Ginger (2012) suggests that this may not be the case for smaller structures and advocates for shorter averaging period in the Australian/New Zealand standard (AS/NZS 1170.2).

The 3-second peak gust speed is used by many standards (Holmes, 2018), including the South African standard (SANS 10160-3:2019). Alternatively, to simplify the dynamic design of structures, the Eurocode (EN 1991-1-4) specifies v_k as a 10-minute mean value that can be converted to a peak gust. This is reasonable over much of western and northern Europe where the strong wind climate is dominated by large synoptic systems which blow consistently and whose fluctuations are approximately Gaussian (Holmes,

2018). A 10-minute mean would not adequately capture the peak gust in regions where storms with a shorter duration, like thunderstorms, dominate the wind climate. This includes parts of South Africa, Australia, and the United States which is why a shorter averaging period is used.

2.4.3 Return Period

A network of anemometers continuously recording wind speeds yields information that can be used to derive v_k. Early loading stipulations in South Africa used the highest wind speed recorded by the nearest anemometer (Goliger *et al.*, 2017). This would result in longer records leading to higher wind speeds which cannot be physically justified. To be more consistent, v_k is defined at a set probability of exceedance p_k and statistical techniques are used to infer it from the available data.

The wind speed corresponding to an average exceedance of once over a 50-year period ($p_k = 0.02$), related to the typical design life of a structure, is commonly accepted as the definition of v_k (Holmes, 2018). This average exceedance period t is referred to as a return period and is related to an exceedance probability p as $t = 1/p$. To estimate v_k requires the relationship between p and measured wind speeds v to be defined using a probability distribution. Numerous approaches and distributions are available (Section 2.3).

2.4.4 Basic Wind Speeds in the South African Standard

The v_k values in South Africa are given in a map (Figure 2.10) and table of characteristic basic wind speeds in SANS 10160-3:2019. These specify one of four $v_k \in \{32, 36, 40, 44\}$ m/s values per metropolitan and district municipality (administrative division of a South African province).

The v_k values in SANS 10160-3:2019 are based on an extensive set of studies (Kruger *et al.*, 2010, 2011, 2013*a*,*b*, 2017). SANS 10160-3:2019 is the first standard to use this map, prior to which the map in Figure 2.11 was used. The older map is based on another set of studies done in the 1980s (Milford, 1985*b*, 1987*a*,*b*), which used OLS to fit a Gumbel

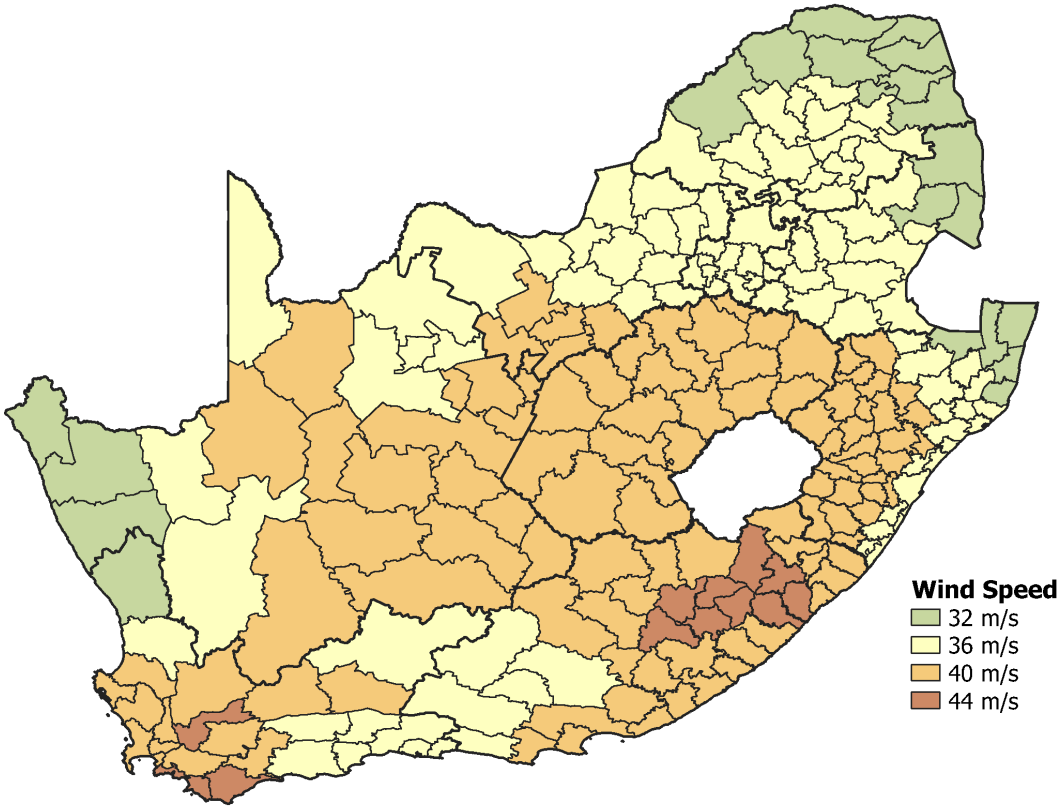


Figure 2.10: Current map in SANS 10160-3:2019.

distribution (Section 2.3), to measurements made with a Dines pressure tube anemometer, between 1948 and 1983, at 15 weather offices throughout South Africa.

The newer map gives more detail because Kruger (2011) used data from more stations (74 versus 15), and could therefore provide new information in many areas. The decision to specify values by municipality decreases the ambiguity of the map and can be physically justified because municipal boundaries often follow topographical features that could correspond to different climatic zones (Kruger *et al.*, 2017).

Kruger (2011) extrapolated to a 50-year return period wind using a Gumbel distribution that was fit using MOM. Kruger (2011) used data recorded with an RM Young anemometer between 1990 and 2008. Unlike Milford (1985*b*), Kruger (2011) accounted for different storm types by separately fitting to each storm type and using the joint distribution, as recommend by Gomes and Vickery (1978), described in Section 2.3.2.

All stations used by Kruger (2011) had at least 10 years of recordings, based on a recommendation by Cook (1985), but the longest was only 16 years. This is less than

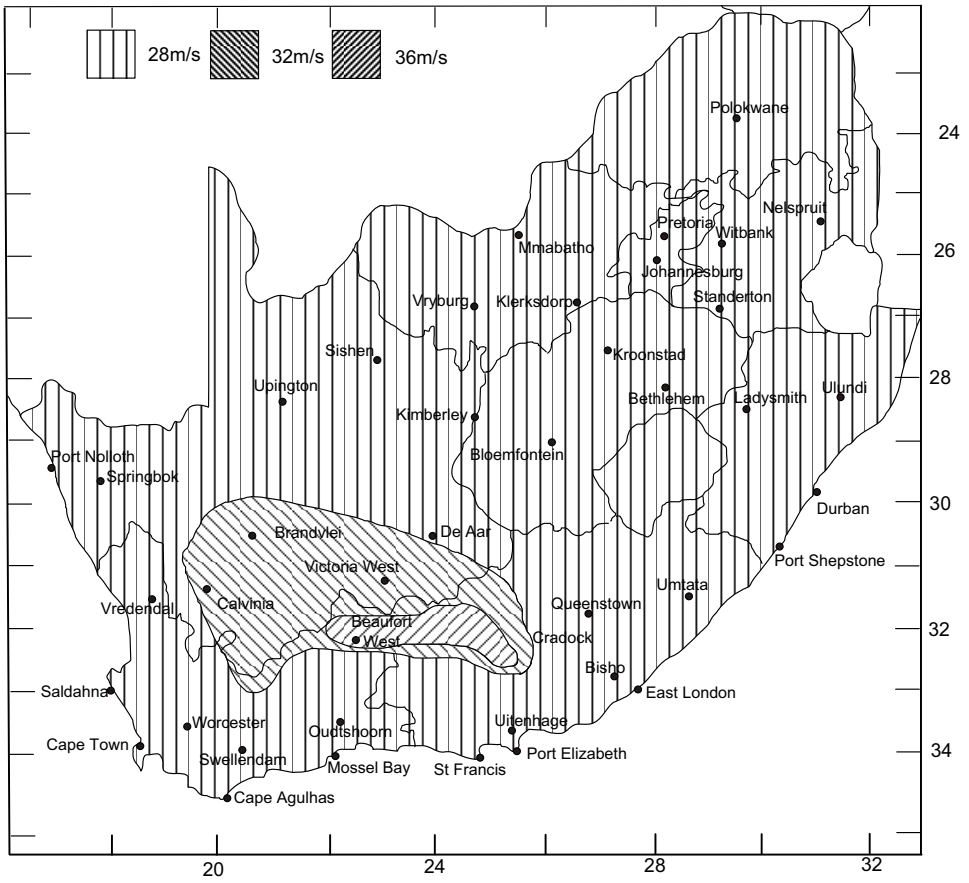


Figure 2.11: Map used prior to SANS 10160-3:2019, note: values were converted to peak gusts using a factor of 1.4.

Milford (1985*b*)'s analysis where up to 35 years were used. Milford (1985*b*)'s data is still available in the SAWS database, which means that for some stations around 70 years of data is available. Unfortunately, a significant discontinuity at the changeover from the Dines pressure tube to RM Young anemometers led Kruger (2011) to reject all Dines data.

Cook (1985) recommend at least 20 years of data for reliable results using a Gumbel analysis, a requirement which none of Kruger (2011)'s data could meet. This shortage led Kruger (2011) to also apply a Peak-over-thresholds (POT) analysis to incorporate more data. Other than the typical drawback of ambiguity in threshold selection in a POT analysis, the inclusion of more data meant that the high standard of quality control that Kruger (2011) applied to the extraction of annual maxima from the SAWS database could not be maintained (Kruger, 2018). Uncertainties in the application of the POT analysis and the limited sample size used for the Gumbel analyses, led Kruger (2011) to

conservatively use the method that yielded the higher extrapolated value and to further adjust the extrapolated value upwards by a 75 % confidence interval.

The incorporation of data collected in the ten years subsequent to Kruger (2011)'s study has been identified as necessary (Goliger, 2016; Goliger *et al.*, 2017). This can also be justified by the simulation study in Section 2.3.5, which showed that significant reduction in error could be achieved by increasing the sample size from 10 to 20.

2.5 Structural Reliability

The fundamental principle in structural design is to ensure that the loading e is greater than the resistance r so that failure, that could lead to loss of life and property, is avoided. This is enabled by modelling the components that make up e and r. These models can never perfectly capture reality and so there is always uncertainty as to how good (near to reality) a model predictions is.

This uncertainty is accommodated with conservatism. A traditional approach to ensure sufficient conservatism is to require that a resistance r_k exceeds a loading e_k by an adequate safety factor $\Gamma = r_k/e_k$, then by specifying a high enough Γ the design should be sufficient. The drawback of this approach is that it fails to consider the variability of individual components in the design. As a result, it is unclear if conservatism has been misassigned.

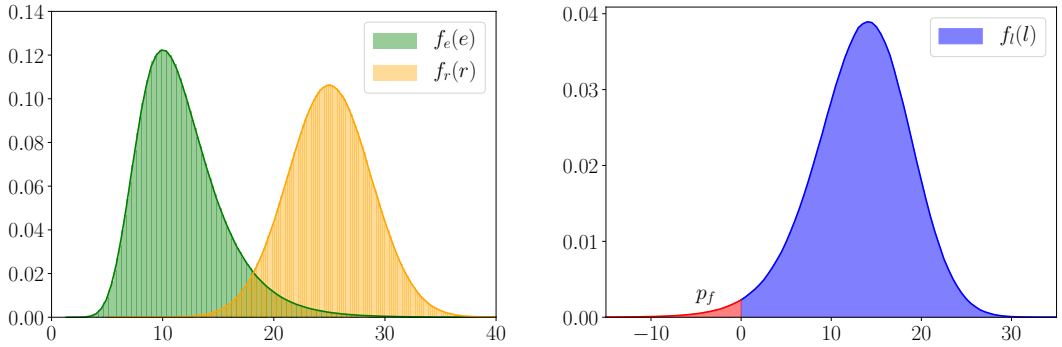
A reliability-based design starts from an acceptable target probability of failure p_t and aims to assign conservatism so that p_t is achieved. This is done by considering a performance function l given by

$$l = r - e, \tag{2.34}$$

where r and e are considered as random variables. Thus, l is also a random variable and the probability of failure $p_f = P(l < 0)$ can be determined.

As long as $p_f \leq p_t$ over the life of the structure, the design can be considered acceptably reliable. The p_f over the t'-year service life of a structure can be related to an annual probability of failure as $(1 - p_{f:life}) = (1 - p_{f:1})^{t'}$. The p_f for a distribution of l, e and r

is shown on probability density functions in Figure 2.12.



(a) Load $f_e(e)$ and resistance $f_r(r)$. (b) Performance function $f_l(l)$.

Figure 2.12: The probability of failure p_f for a distribution of load e and resistance r.

2.5.1 Reliability Calibration of Wind Loading Standards

The t-year return period wind speed has a high probability of exceedance over the life of the structure. If the typical $t = 50$ is used, then $p_{f:life} = 1 - (1 - 1/50)^{50} = 0.64$, which would be unacceptable as a final design value. Therefore, a characteristic load $q_k = \rho c_k v_k^2/2$ is adjusted to a design value q_d with an acceptable probability using a partial safety factor $\gamma_w = q_d/q_k$.

Modern standards use a reliability assessment to ensure that γ_w is acceptable. A suitable performance function for the reliability assessment of a wind loading standard is

$$l = r - (g + q). \tag{2.35}$$

Where r, g and q are random variables representing the structural resistance, permanent loading and wind loading respectively. To enable generalisation, the variables in Eq. 2.35 can be normalised by their design values r_d, g_d, and q_d

$$l = \frac{r}{r_d} - (1 - \eta)\frac{g}{g_d} - \eta\frac{q}{q_d}. \tag{2.36}$$

To consider a wide variety of design scenarios, different ratios of wind to total load $\eta = q_d/(q_d + g_d)$ can be used.

If it is assumed that the design value is equal to the product of its partial factor and characteristic value, then

$$l = \frac{r}{\gamma_r r_k} - (1-\eta)\frac{g}{\gamma_g g_k} - \eta\frac{cv^2}{\gamma_w c_k v_k^2}, \tag{2.37}$$

where the random wind load is represented as the product of the squared wind speed v^2 and other modification factors c. This formulation allows variables to be specified relative to their characteristic values and for the relevant partial safety factors γ_r, γ_g and γ_w to be considered explicitly.

The reliability performance of the standard can then be assessed by calculating the p_f using a suitable reliability method (Section 2.5.2). This requires the distributions of the random variables relative to their characteristic values $\{r/r_k, g/g_k, c/c_k, v/v_k\}$ and partial factors $\{\gamma_r, \gamma_g, \gamma_w\}$ to be defined (Retief and Dunaiski, 2009; Botha *et al.*, 2018*b*).

Annual maximum wind speeds v typically have a coefficient of variation σ_v/μ_v of around 0.1-0.35 (JCSS Model Code). The overall variability of c is probably in a similar range (Botha *et al.*, 2018*b*), by squaring the wind speeds to convert to wind pressures the variability is roughly doubled. Because these variabilities approximately combine as sums of squares, the wind speed variability typically dominates the uncertainties and contributes around 70 % to 80 % towards the overall uncertainty of the wind load (Cook *et al.*, 2011). Therefore, it is important that variability in v is appropriately represented in the standard.

2.5.2 Computation of the Probability of Failure using FORM

The p_f can be related to the probability density of the performance function $f_l(l)$ as

$$p_f = \int_{-\infty}^{0} f_l(l)dl. \tag{2.38}$$

Holický (2009) describe a number of approaches to the computation of p_f. In some simple cases the integral above can be computed analytically, or for less simple scenarios numerical integration could be applied (Gulvanessian *et al.*, 2012). Commonly l involves too many operations or more complicated probability distributions and numerical integration is unfeasible. In this case other methods, such as Monte Carlo simulation, need to be applied. In many situations approximate methods offer a more efficient solution, one of the most widely applied of which is the first-order reliability method (FORM).

To carry out a FORM analysis the distribution of each variable is transformed to an equivalent standard normal distribution, where each variable x_i is transformed to a standard normal variable z_i such that

$$\Phi(z_i) = F_{x_i}(x_i), \tag{2.39}$$

where $F_{x_i}(x_i)$ and $\Phi(z_i)$ are the original and standard normal cumulative density functions respectively. A realisation of the random variables in the performance function $\{x_1, x_2 \dots x_m\}$ can be transformed to a standard normal vector $\mathbf{z} = \{z_1, z_2 \dots z_m\}$ and visualised as a point in m-dimensional standard normal space (Figure 2.13).

A limit state function $l(\mathbf{z}) = 0$ divides this space between failure and non-failure. The point on this limit state nearest the origin $\mathbf{z}_d$ is the *design point*, which can be interpreted as the most likely combination of values that result in failure. FORM approximates $l(\mathbf{z}) = 0$ by its tangent at the design point. The minimum distance β of this tangent to the origin is given by

$$\beta = \min(\sqrt{\mathbf{z}^T\mathbf{z}}), \tag{2.40}$$

and is normally a good approximation of the probability of failure $p_f = 1 - \Phi(\beta)$, provided the limit state is not highly non-linear in standard normal space. The β can be found using a numerical optimisation procedure (Hasofer and Lind, 1974; Rackwitz and Flessler, 1978).

The direction cosines of the limit state at $\mathbf{z}_d$ are known as sensitivity factors α, for

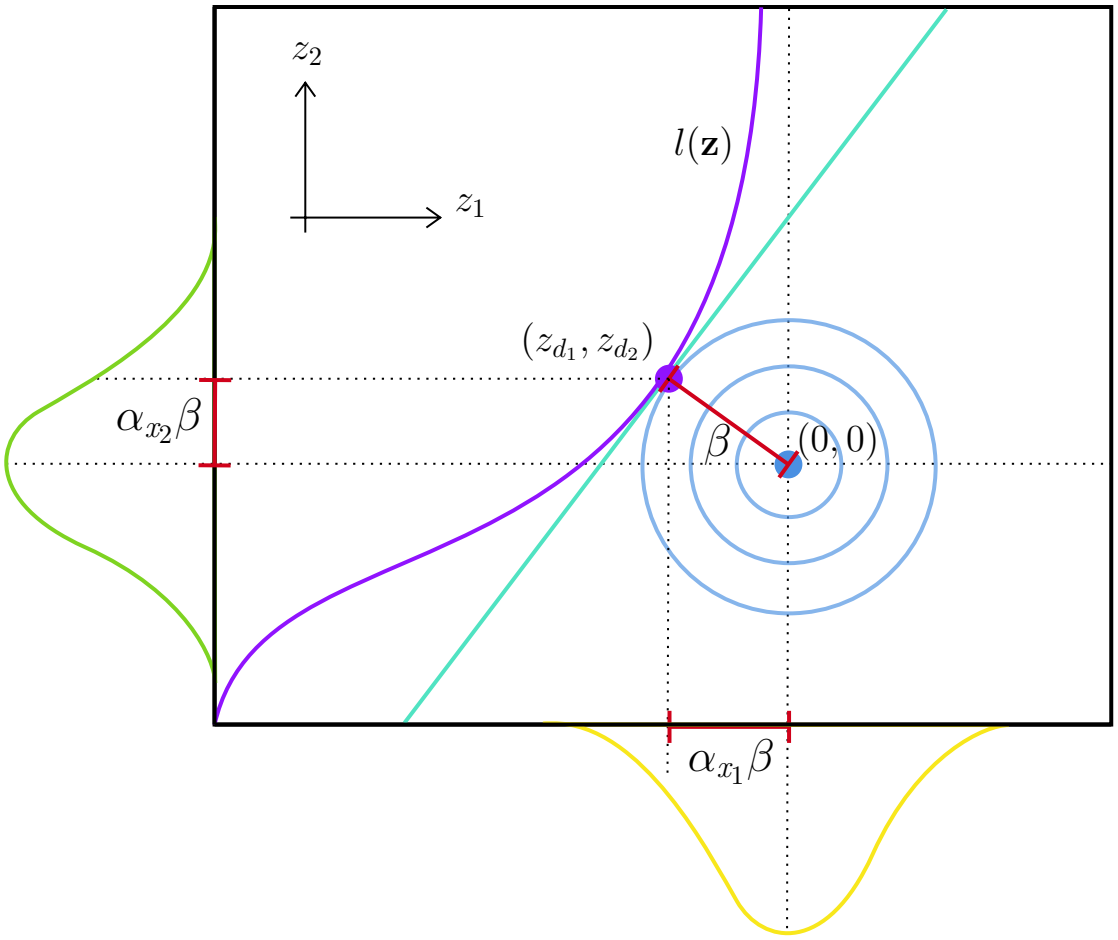


Figure 2.13: FORM analysis in two dimensions

example the sensitivity factor for x_1 is given by

$$\alpha_{x_1} = \frac{\partial l(\mathbf{z}_d)/\partial z_1}{\sqrt{\sum \left[\partial l(\mathbf{z}_d)/\partial z_i\right]^2}} \tag{2.41}$$

with

$$\sum \alpha_{x_i}^2 = 1. \tag{2.42}$$

These factors provide a measure of the relative contribution of a variable to the overall uncertainty of the formulation and are a useful tool for identifying where efforts to reduce uncertainties could be focussed.

The elements of a FORM analysis with two random variables are shown in Figure 2.13.

2.5.3 Reliability of the South African Standard

Botha *et al.* (2018*a*) performed the most recent reliability analysis of the South African standard. This led to an increase of γ_w from the previous 1.3 to the current 1.6 in SANS 10160-3:2019.

The distributions and partial factors of a representative permanent loading (g/g_k,γ_g) and four different representative resistances (r/r_k,γ_r), which cover a range of materials, recommended by Retief and Dunaiski (2009) were used. The distributions of c/c_k and v^2/v_k^2 relevant to South Africa were defined by Botha (2016). Botha *et al.* (2018*a*) considered different partial factors and calculated the p_f using FORM. A $\gamma_w = 1.6$ was found to be adequate given the range of resistance r, load ratios η, and wind load modification factors c considered. The dominance of the wind speed towards the total uncertainties were also confirmed by the FORM sensitivity factors (Botha, 2016).

A potential issue with Botha *et al.* (2018*a*)'s calibration is that a single distribution of v^2/v_k^2 was used, when in reality the distribution could differ from site to site (Ellingwood and Tekie, 1999). This may be acceptable if the distributions are reasonably similar (Hosking and Wallis, 2005), as is the case over much of Europe (Holmes, 2018).

In the United States and Australia/New Zealand the geographical variation of the wind climate is significant, most notably between regions that are dominated by tropical cyclones and those that are not. In these cases the distribution of v varies significantly by region and so using a single partial factor may lead to significant geographical variation of the reliability level of the standard. Instead, wind loading standards of these countries (ASCE 7-16; AS/NZS 1170.2) have incorporated γ_w into a high return period wind speed which achieves the target reliability. This should lead to more uniform reliability performance (Hong *et al.*, 2016) throughout the area covered by the standard.

The South African extreme wind climate is not significantly affected by tropical cyclones (Section 2.2) and thus, should be less diverse than the wind climate of the United States or Australia/New Zealand. Nevertheless, the dominance of synoptic and thunderstorms in different regions means that the reliability level could still differ regionally, and a more direct consideration of design values could be considered.

2.6 Conclusion

The South African extreme wind climate is dominated by synoptic winds in the south-west and thunderstorms in the north-east. These winds are deemed the most appropriate to focus on for the South African wind loading standard. Other less frequent types of extreme winds caused by tropical cyclones or tornadoes do not appear to warrant specific consideration at this stage.

Extreme value theory provides statistical techniques to characterise the winds from these events. Using the block maxima approach, only the maximum value per year is extracted from the record and an asymptotic distribution is fit to characterise these maximums. This has the advantage of simplicity but can lead to high model-variance due to the relatively small quantity of data used. Of the asymptotic forms, the Gumbel distribution suffers the least from this drawback, although significant model-bias can be introduced.

The Gumbel distribution can be fit using a number of techniques, including the method of moments, graphical methods, and maximum likelihood estimation. Maximum likelihood estimation and the graphical method, fit using generalised least squares, led to the lowest average error in a Monte-Carlo simulation investigation. This also demonstrated that increasing the sample size was an effective way of reducing error, and how principles of model selection could be applied to inform the characterisation of available data.

The South African standard uses a fairly common format to prescribe wind loads, where the severity of the wind climate is represented by a characteristic basic wind speed, with an annual exceedance probability of 0.02, that is specified by municipality. Incorporating recently collected data into the standard was identified as a need that this research could fulfil.

An overview of how wind speed uncertainties, which typically dominate, are incorporated into the partial factor was also presented. This included a summary of reliability techniques.

Chapter 3

Dataset Creation

Summary

Climatic data is the primary input to any potential model for the estimation of design wind loads in a standard. Maximising the quantity of relevant high quality data available allows more nuanced, and thus potentially more accurate, models to be investigated. To this end, an extensive dataset of annual maximum wind gusts was created for refinement of the South African wind loading standard. This new dataset is built upon the dataset used for the previous iteration of the standard with additional data extracted from the South African Weather Service database. These gusts were quality controlled and normalised by correcting for non-standard surface roughness using a Geographic Information System approach. An attempt to incorporate previously rejected historic data, recorded using Dines anemometers, was made. This involved the implementation of a technique developed since the previous update, that uses random process and linear system theory to account for the amplitude response of the different instruments. The result of these efforts is an updated dataset of 3500 annual gust maxima, recorded at 132 stations throughout South Africa, with an average recording period of 22 years, representing both an improved spatial resolution and increased series lengths of available data.

Notation

v	Wind speed
c_r	Roughness correction factor
c_h	Height correction factor
z	Height
z_o	Roughness Length
$\overline{v}$	Average wind speed
I_v	Turbulence intensity
d	Instrument distance constant
H_D	Amplitude response of dines instrument
H_{aws}	Amplitude response of AWS instrument
G	Gust factor
g	Peak gust factor
σ_v	Standard deviation of wind speeds
T	Averaging time
$S_v(f)$	Spectral density function of the wind speed fluctuations
ν	Cycling rate
$H(f)$	Amplitude response
v_{aws}	AWS wind speed measurement
v_D	Dines wind speed measurement
p_A	P-Value of Anderson Darling test

3.1 Introduction

To review the representation of the wind climate in SANS 10160-3:2019, wind speeds with the potential to cause structural damage throughout South Africa are studied. These studies are founded on historical wind speed observations that should span as long a period at as many locations as possible. To enable generalisation, data should reflect standard conditions and not be specific to a particular location or measurement instrument. A dataset which aims to fulfil these requirements was created. This involved extraction of relevant data from a central climatic database (Section 3.2), normalisation of site conditions (Section 3.3), and normalisation of instrumentation response (Section 3.4).

An earlier version of the work detailed in this investigation has been documented in Bakker and Viljoen (2019). The main refinements since this earlier version include implementation of World Meteorological Organization (2008) suggestions in the surface roughness correction, the conservative application of surface roughness correction to thunderstorm gusts, and a more sophisticated instrumentation correction procedure in conjunction with specific consideration of the dynamic response of the anemometers used

by the South African Weather Service.

3.2 Data Extraction

3.2.1 Sources of Data

The largest source of available wind speed data in South Africa is managed by the South African Weather Service (SAWS). The SAWS maintains a database with records from an extensive network of 231 operational Automatic Weather Stations (AWS) and historical data from a further 156 inactive stations, the network is shown in Figure 3.1.

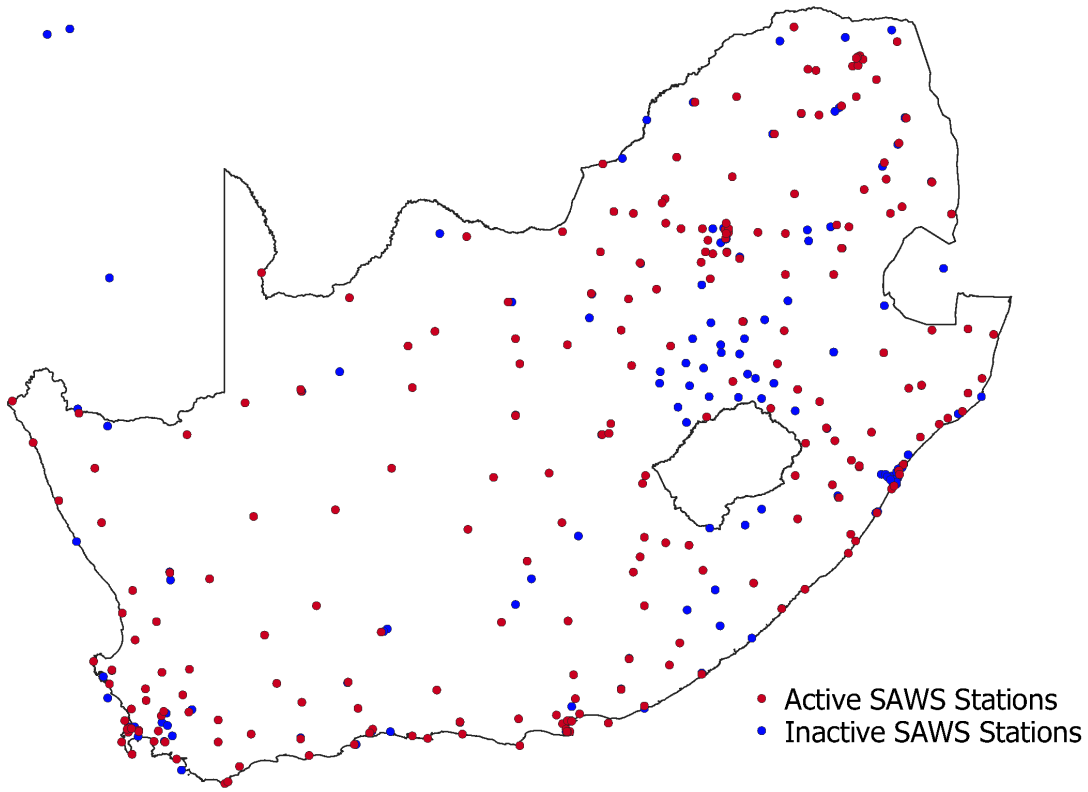


Figure 3.1: Station locations in the South African Weather Service database.

The SAWS network has seen substantial growth over time, as shown in Figure 3.2, with a sharp increase around 1990. Before 1990 stations were manned, which was expensive and restricted the size of the network. From 1990 less expensive AWS started to be deployed, which explains the increase in network growth at this time. Measurements

of mean wind speed, temperature, pressure, humidity, rainfall, wind direction and the maximum 3-second wind gust, typically at a five minute resolution, are available at these stations. The measurements made by the SAWS aims to meet World Meteorological Organization (2008) standards.

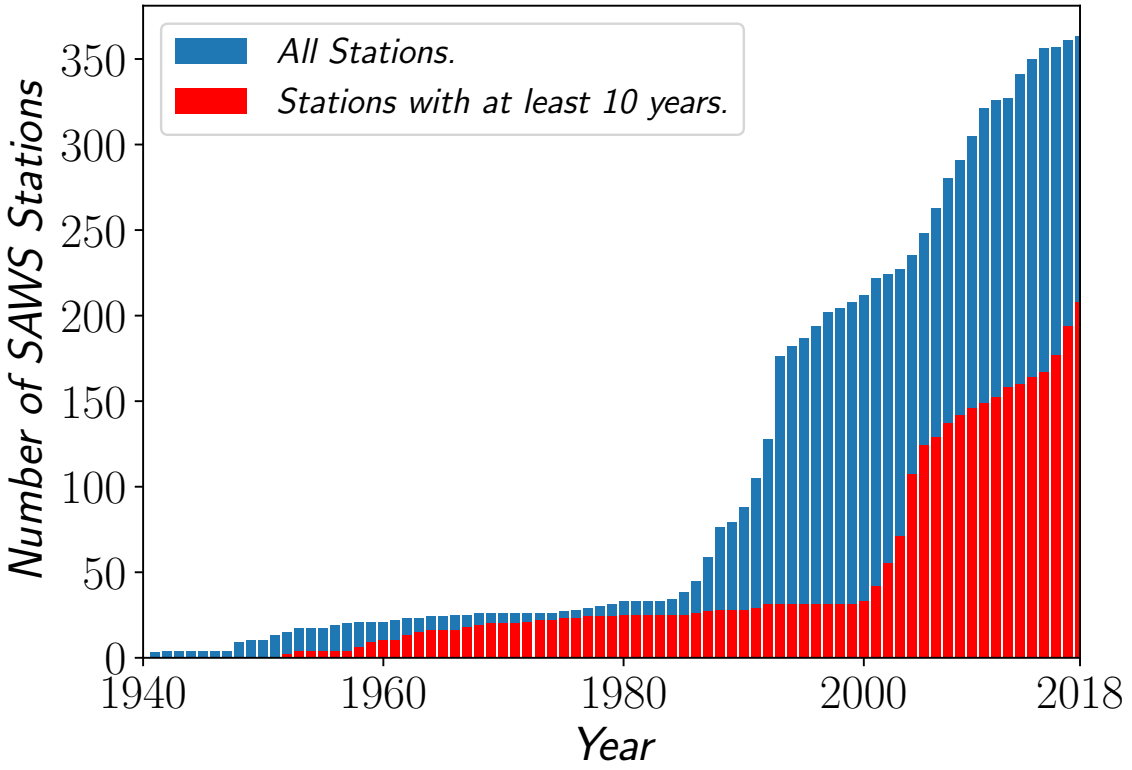


Figure 3.2: Number of stations in the South African Weather Service database over time.

The SAWS database includes an estimated 150 million gust recordings. The majority of which are irrelevant to SANS 10160-3:2019 as they are caused by wind generating mechanisms unlikely to cause damage. A simple systematic approach to finding the relevant gusts is needed. The traditional solution is to only consider the annual maximum gusts (Palutikof *et al.*, 1999). This approach was followed by Kruger (2011), who extracted annual maxima at 92 stations, throughout South Africa, from the SAWS database. Kruger (2011) used this data to characterise the extreme wind environment of South Africa, which served as the foundation for the current characteristic wind speed map in SANS 10160-3:2019. Kruger (2011)'s dataset, of around 1500 extensively quality controlled gusts, was

considered the most practical starting point for an expanded dataset.

Supplementing Kruger (2011)'s data with the subsequent 10 years is the obvious first step in extending the dataset. Following recommendations by Cook (1985), Kruger (2011) only considered stations for which at least 10 years, each with less than 10 % missing data, were available. Since Kruger (2011)'s study, another 40 stations now meet this requirement and can also be considered for inclusion into the extended database. The spatial distribution of these 40 stations along with the 92 used by Kruger (2011) is shown in Figure 3.11.

Kruger (2011) rejected data recorded before 1990 which used a Dines anemometer from his dataset. This instrument had been equipped at major (first order) weather stations by the South African Weather Bureau (predecessor to the SAWS). Some of these stations have records that extend back to the 1940s, and could provide long series which would substantially reduce uncertainties (Palutikof *et al.*, 1999; Rózsás and Sỳkora, 2016). Additionally, these stations, despite making up a small proportion of the total, are of higher relevance to the built environment as they are typically situated at airports nearer larger urban centres (Figure 3.11). The possibility of substantially reducing uncertainties of wind speed estimates where they are most required led to the re-evaluation of the Dines data (Section 3.4).

The SAWS database contains recordings from around 25 locations throughout South Africa prior to the implementation of the AWS network. Only stations where an active AWS station is still in operation were considered. The justification being that AWS data could be used for verification purposes and that these stations were likely to have the longest records. The climatic data required for separation of gusts by climatic mechanism (Section 3.2.2) is unavailable at Dines stations. Therefore, only stations where the corresponding AWS data indicated dominance of a single climatic mechanism were considered.

These restrictions leave twelve stations (Table 3.1) which cover the majority of metropolitan areas throughout South Africa. Ten of these stations were used by Milford (1987*b*) to determine the characteristic wind speeds prior to the current SANS 10160-3:2019 version.

The other two stations (Irene & Langebaan) were not used by Milford (1987*b*), presumably because their records were too short at the time.

Station	Year Established	Anemometer Height
Alexanderbaai	1951	12m
Bloemfontein	1962	12m
Cape Town	1956	12m
Durban	1956	13m
East London	1948	14m
Irene (Pretoria)	1976	15m
Johannesburg	1953	13m
Kimberly	1941	12m
Langebaan	1975	13m
Polokwane (Pietersburg)	1951	12m
Port Elizabeth	1949	15m
Upington	1970	12m

Table 3.1: Dines stations considered.

3.2.2 Separation by Climatic Mechanism

Most statistical distributions assume that data is independent and identically distributed. While independence of annual maxima is typically not an issue, gusts that originate from different causes are likely to be distributed differently. In order to apply conventional statistical distributions, they should be fit to wind speeds originating from different causes separately and then the joint distribution be considered to predict the probability of exceedance of a given wind speed (Gomes and Vickery, 1978).

Holmes (2018) identifies four major causes of winds with the potential to inflict significant damage: extra-tropical cyclones, tropical cyclones, thunderstorm downbursts and tornadoes. These are discussed in Section 2.2. While all four of these mechanisms have been observed at some point in South Africa, tornadoes and tropical cyclones are unlikely to cause damage on a scale that warrants consideration in the loading standard (Kruger, 2011).

Kruger (2011) performed a detailed study on the cause of each annual extreme in his dataset and identified six different causes of annual maxima. Kruger (2011)'s breakdown

by cause is given in Figure 3.3 which shows that the majority, around 86%, of annual maxima can be attributed to just two primary causes: thunderstorms (downbursts) and cold-fronts (extra-tropical cyclones). The other four secondary causes account for a small proportion of the total events. These secondary causes are difficult to distinguish from one another without synoptic charts and meteorological knowledge. Specific analysis of these secondary causes is unlikely to have a significant overall impact due to the small proportion of the overall dataset they make up. To simplify the process, secondary mechanisms were grouped along with cold fronts as synoptic events, as is typically done (Gomes and Vickery, 1978; Holmes, 2018; Cook *et al.*, 2003), in the extended dataset.

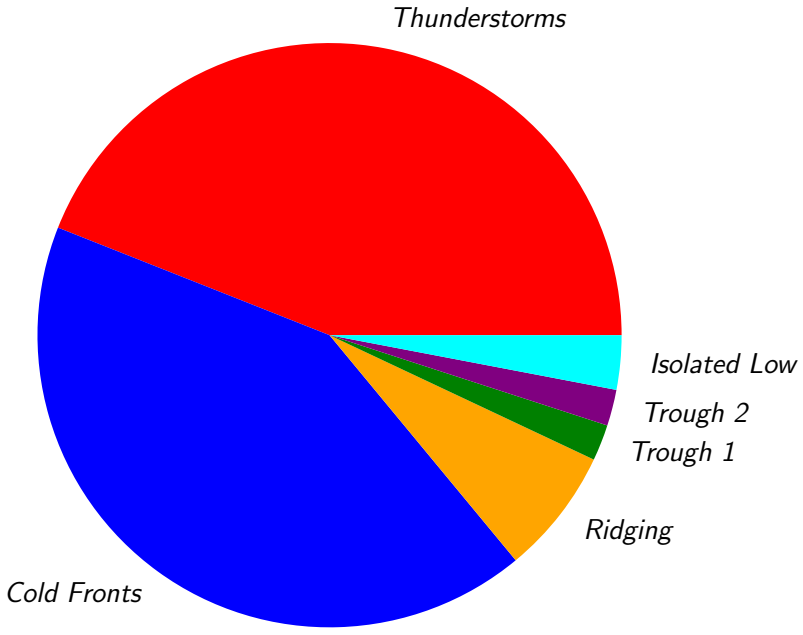


Figure 3.3: Proportortion of gusts by cause, classified by Kruger (2011).

As annual maxima were extracted from the SAWS database, they were identified as thunderstorm or non-thunderstorm/synoptic. Kruger (2011) visually inspected the

5-minute resolution climatic data to identify thunderstorm events. These typically display a number of characteristics including a high ratio of gust to 5-minute mean wind speeds, the commencement of rain or hail, a temperature decrease, a pressure and humidity increase, and a shift in the wind direction. To streamline thunderstorm identification when extending Kruger (2011)'s data, a script was written to search through the database and identify potential thunderstorms, which were then confirmed visually before inclusion in the extended dataset. Further study into full automation of this labour intensive task would be useful to help process the ever increasing quantities of data that are becoming available.

Stations dominated by thunderstorms are generally found in the north-east half of the country and those dominated by synoptic conditions are found in the south-west. Mixed stations, whose annual maxima demonstrate both mechanisms, are found between these zones (Figure 3.11). These mixed stations yield two maximum gusts per year, one thunderstorm and one synoptic.

3.2.3 Removal of Erroneous Measurements

While segregating mechanisms, spurious spikes, often of extremely high magnitude (>70 m/s), were observed in the SAWS wind record. These were not accompanied by any corresponding climatic information from pressure, humidity or rainfall sensors which could explain them. The same phenomenon was observed by Kruger (2011) and attributed to instrumentation malfunction.

The SAWS has a quality management system in place that aims to ensure that data collected by AWS are accurate. Potentially erroneous gusts are automatically flagged when recorded and designated data quality controllers are responsible for reviewing these gusts and removing erroneous measurements from the record (South African Weather Service, 2017). Unfortunately this system does not catch all erroneous measurements (Kruger, 2018). Therefore, to ensure accuracy of the extended dataset, the climatic history was visually inspected and gusts were rejected if they could not be justified.

Around 10% of annual maxima recorded in the SAWS database between 2008 and

2018 were found to be erroneous and removed. The proportion of erroneous measurements varied significantly by station and could make up a large proportion of recorded annual maxima. For example at the East London Weather Office, shown in Figure 3.4, around 40% of annual maxima were removed.

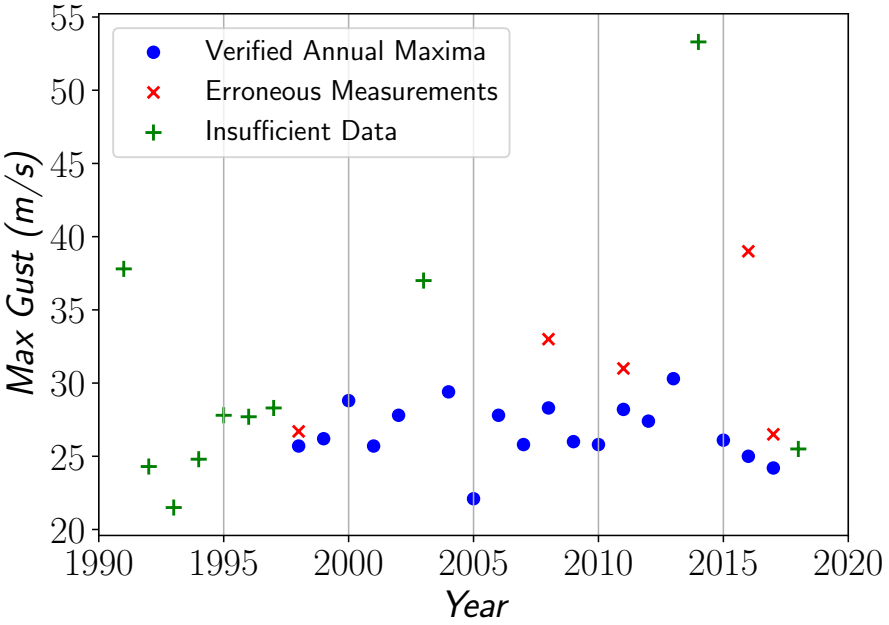


Figure 3.4: Annual maxima for the East London Weather Office.

Verification of pre-2008 data, originating from Kruger (2011)'s dataset, was unnecessary as they had already been verified by Kruger (2011). Similar to segregation of thunderstorms, verification of the data proved to be laborious.

Often there was insufficient data for a given year to warrant inclusion in the extended dataset. An annual maximum was only considered if the AWS had been operational for at least 90% of that year. Lower resolution AWS data recorded before 1995 could not be verified and so were not included in the expanded dataset. Data recorded before 1990 could also not be verified, but because it was collected manually and the volume of data at that time was far less, it was subjected to more stringently quality control than AWS data (Kruger, 2018). Therefore, these recordings were not rejected from the updated dataset at this stage.

3.3 Surface Roughness Normalisation

The regional wind speed in SANS 10160-3:2019 refers to 3-second gust wind speed, measured at 10 metres above ground level in open country terrain with low vegetation. This aligns with World Meteorological Organization (2008) standards, to which the SAWS aims to adhere. Gusts measured under different conditions should be corrected to these standard conditions to be appropriate as input to SANS 10160-3:2019.

3.3.1 WMO Correction Method

The World Meteorological Organization (2008) recognise that it is practically difficult for stations to fully comply with their requirements and provide guidance for the correction of non-standard conditions.

Where topographical features or obstructions are likely to significantly impact measurements the World Meteorological Organization (2008) recommend wind tunnel testing for correction. This is obviously infeasible in the current study given the number of stations for which it would be required. Stations with these issues were noted so that if it is found that they lead to a significant difference in results, the decision to include them can be re-evaluated.

According to the World Meteorological Organization (2008), provided flow distortion and topography problems are negligible, non standard conditions can be accounted for as

$$v_c = c_r c_h v. \tag{3.1}$$

Where the measured wind speed v is corrected using c_r, which accounts for non standard terrain roughness as

$$c_r = \frac{\ln(60/z_o)\ln(10/0.03)}{\ln(10/z_o)\ln(60/0.03)}, \tag{3.2}$$

and c_h which accounts for a non standard instrument height z as

$$c_h = \frac{\ln(10/z_o)}{\ln(z/z_o)}. \tag{3.3}$$

The roughness length z_o is the height at which frictional effects lead to a wind speed of zero, given a logarithmic boundary layer profile (Kruger, 2011). Certain types of land cover have higher friction which result in a greater z_o. The z_o for a number of land cover classes is specified by the World Meteorological Organization (2008), given in Table 3.2, where $z_o = 0.03$ metres corresponds to World Meteorological Organization (2008) standard conditions.

Table 3.2: Roughness length by terrain type.

Terrain Description	z_o (m)
Open sea	0.000 2
No vegetation, no obstacles: mud flats, snow	0.005
Open flat terrain: grass, few isolated obstacles	0.03
Low crops, occasional large obstacles	0.10
High crops, scattered obstacles	0.25
Parkland, bushes, numerous obstacles	0.5
Regular large obstacle coverage (suburb, forest)	1.0
City centre with high and low-rise buildings	2

3.3.2 Determination of Correction Factors

While the instrument height z of SAWS anemometers conforms to the 10 metre requirement ($c_h = 1$), the surrounding terrain roughness can deviate substantially from World Meteorological Organization (2008) standard conditions (Kruger, 2011). Therefore, c_r should be determined to normalise measurements. This requires a representative z_o for each measurement, which is difficult to determine as z_o typically varies both by station and by wind direction at that station. The most accurate method of determining z_o, recommended by the World Meteorological Organization (2008), requires higher resolution wind speed recordings than is available in the SAWS database.

Alternatively, the World Meteorological Organization (2008) recommend that a visual survey of terrain conditions for two kilometres upwind of a station be used to assign a z_o to each 30° wind direction sector. While the distance of surface roughness influence will vary depending on wind speed, surface roughness, and instrument height (Wieringa,

1992), the World Meteorological Organization (2008) two kilometre recommendation is considered simple enough for widespread implementation and reasonable for an anemometer positioned at 10 metres above ground level.

This was essentially the method used by Kruger (2011) to correct winds in his dataset. Performing a visual inspection for each of the twelve direction sectors of each station would be both subjective and labour intensive. Instead in this study, z_o values were determined automatically using a Geographical Information System (GIS).

The South African Department of Environmental Affairs has classified the entire country into 72 land cover categories at a 30m×30m resolution. This data is available for 1990, 2013/2014 and 2018 to download as a raster layer (Department of Environmental Affairs Republic of South Africa, 2020). The 2013/2014 layer was selected as the most representative layer for the period of recordings. The land cover for the area surrounding the Struisbaai station is shown in Figure 3.5.

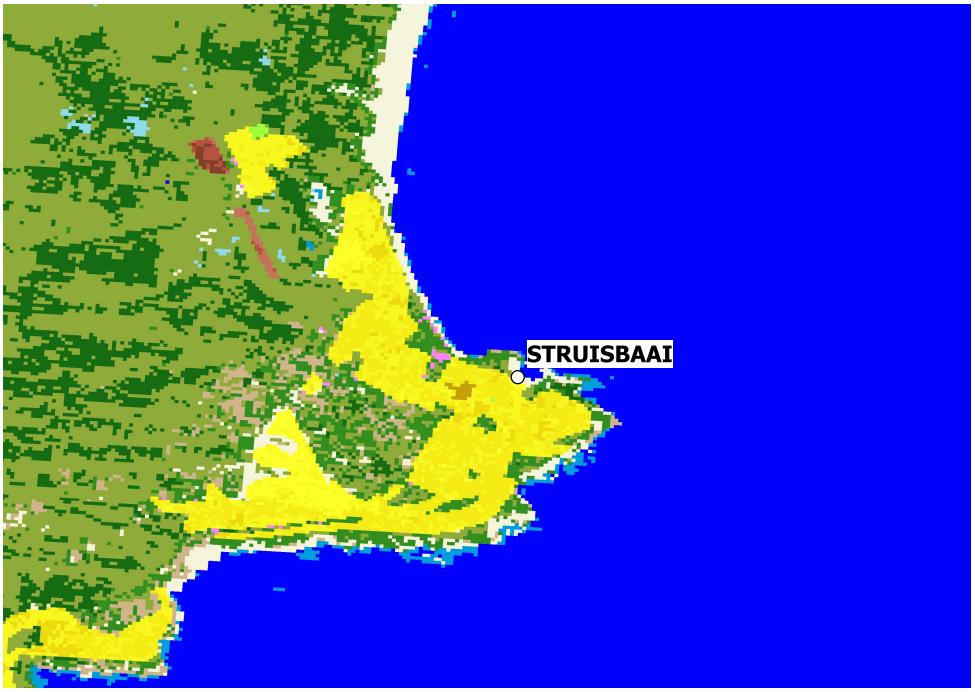


Figure 3.5: Land-cover surrounding Struisbaai AWS (Legend to land-cover classes is given in Table 3.A1).

A z_o was assigned to each of the land cover categories using Table 3.2. The land cover GIS layer was then transformed to a z_o layer based on these assignments (Table 3.A1).

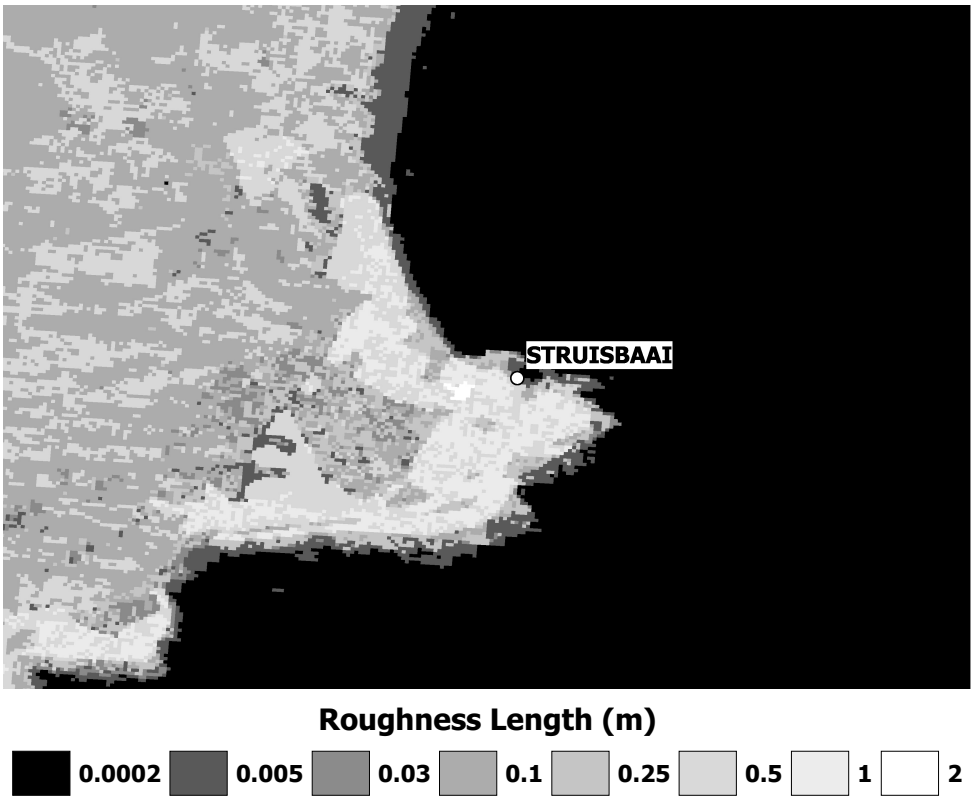


Figure 3.6: Roughness of terrain surrounding Struisbaai AWS.

The resulting z_o layer is shown in Figure 3.6 for the area surrounding the Struisbaai AWS.

The precise location of weather stations was then required. The SAWS database only provides locations to within around 1.5 kilometres, which was insufficient for accurate assessment of roughness conditions. More precise locations were found in station meta-data reports, although these were often incorrect, commonly due to confusion about degrees minutes and seconds versus decimal degrees. This meant station locations (to within a few metres) had to be verified using Google Earth.

The station locations were imported into the GIS system and a two kilometre zone was drawn around each. These zones were then subdivided into 30° sectors and the average z_o was calculated for each sector as

$$z_o = \exp\left[\frac{1}{m}\sum_{i=1}^{m}\ln(z_{oi})\right], \tag{3.4}$$

where m is the number of 30m $\times$ 30m land cover blocks and z_{oi} is the surface roughness

of each block. Aggregating $\ln(z_o)$ rather than z_o is recommended by the World Meteorological Organization (2008). Using these estimated z_o values, correction factors could be calculated with Eq. 3.1 per sector for each station. The resulting factors are shown in Figure 3.7 for the Struisbaai station.

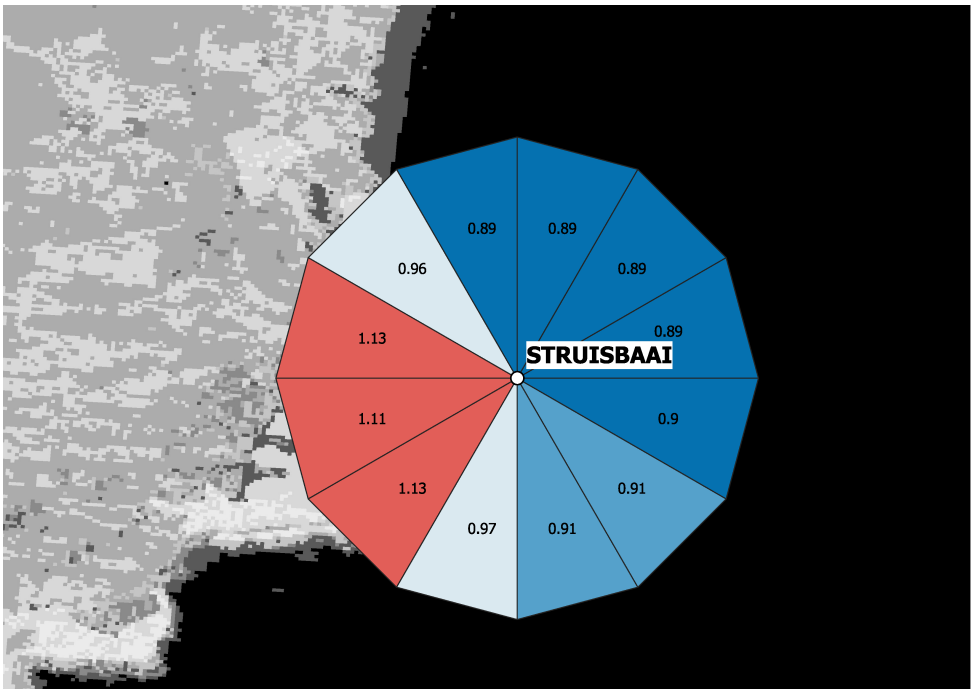


Figure 3.7: Correction factors by 30° direction sector for Struisbaai AWS.

3.3.3 Application of Correction Factors

The World Meteorological Organization (2008) correction factors are derived under the assumption of a well developed boundary layer profile. Synoptic winds are caused by large pressure systems, which flow over the earth's surface for long enough to be affected by surface friction and develop a boundary layer profile. In this case, consideration of z_o for around two kilometres upwind is applicable to an anemometer positioned at 10m above ground level. This may not be the case for thunderstorms which originate from a strong downdraft of air. As a result, the correction factors are potentially invalid for thunderstorm winds (Holmes, 2018).

Near the origin of a downdraft, where wind has not blown over the surface for a

substantial distance, the terrain roughness is unlikely to significantly impact wind speed. In this case, the application of the above correction factors are likely inappropriate. As wind blows further from the origin, it starts to be impacted by the underlying terrain and approaches a typical boundary layer profile (Holmes, 2018). In this case, the application of the correction factors could be considered.

The shape of the wind speed profile, and corresponding correction factor, for a particular thunderstorm gust, is evidently dependant on the unknown distance to the storm origin. This uncertainty and a provision in ISO 4354:2009 led Kruger (2011) to leave thunderstorm gusts uncorrected. If correction factors were applied to thunderstorm winds it was found that 93% would be increased. The application of correction factors is therefore generally conservative.

An issue with leaving thunderstorm gusts uncorrected is that SANS 10160-3:2019 assumes a boundary layer profile and makes no specific provisions for thunderstorm profiles. Designers are thus permitted to adjust loads according to surrounding terrain conditions, based on measurements that did not undergo the same treatment.

Given limited information on more appropriate correction factors and for the sake of consistency with SANS 10160-3:2019, it was decided to opt for the more conservative option and apply the correction factors to both synoptic and thunderstorm gusts. This was achieved by matching the wind direction of the gust to the correction factors derived in Section 3.3.2.

3.4 Instrument Normalisation

The primary motivation for rejection of pre-1990 data by Kruger (2011) was an observed discontinuity when Dines anemometers were replaced by RM-Young anemometers used at AWS. The pre-1990 mean wind speeds were generally lower than post-1990, this could be explained by a left censoring of the data at 1.5 m/s when digitising the pre-1990 records (Kruger, 2011; Goliger, 2016). The focus of the current study is on gust data, which also displayed a discontinuity at the instrumentation changeover. Contrary to mean speeds,

Dines measurements of gust wind speeds were typically higher.

Discontinuities at instrumentation changeover has been an issue in various other studies (Miller *et al.*, 2013; Masters *et al.*, 2010; Department of the Environment, Heritage and Local Government, 2009; Safaei Pirooz and Flay, 2018). Of particular interest in South Africa is the case in Australia, which also made use of the Dines anemometer until the 1990s (Masters *et al.*, 2010). Since Kruger (2011)'s study, an investigation into the dynamic response of the Dines instrument by Miller *et al.* (2013) has been conducted.

Following experimental determination of the frequency response of the Dines instrument (Figure 3.8), Miller *et al.* (2013) used random process and linear system theory to determine a ratio between gust wind speeds recorded using the Dines and AWS instruments. This showed that the different frequency response could largely explain the discrepancies between instrumentation and could be used to account for the observed discontinuity. Adaptation of their methodology is investigated to correct the data from Dines instrumentation in South Africa.

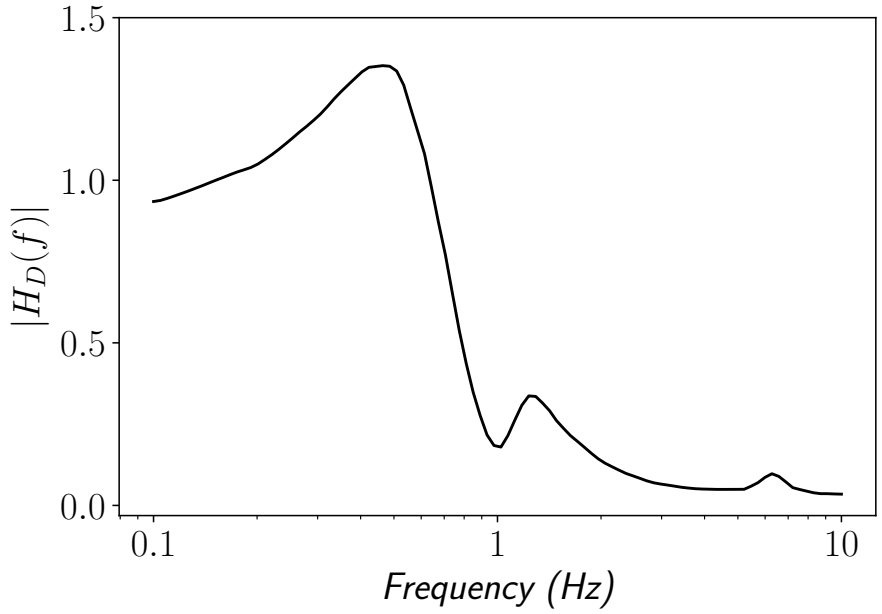


Figure 3.8: Amplitude response of Dines instrument as determined by Miller *et al.* (2013).

3.4.1 Correction Method

The ratios between Dines and AWS instruments calculated by Miller *et al.* (2013) range between 1.06 and 1.12, dependant on mean wind speed $\overline{v}$ and turbulence intensity. Turbulence intensity I_v is the ratio of standard deviation to the mean of the fluctuating wind speed (Holmes, 2018).

The response of an anemometer can be quantified using a distance constant d, which is defined as the distance of air that flows past an anemometer during the time it takes the propeller to reach 63% of equilibrium speed following a change in wind speed (Brock *et al.*, 2001). Miller *et al.* (2013)'s ratios were derived based on the response of Synchrotac 706 cup-anemometers which are used at Australian AWS. In South Africa RM-Young anemometers are used. The d, and hence the response, differs between instruments with $d = 13$m and $d = 2.7$m for the Synchrotac 706 and RM-Young anemometers respectively (Miller *et al.*, 2013; R.M. Young Company, 2005). Therefore, the method described by Miller *et al.* (2013) should be applied specifically considering the RM-Young anemometer.

The amplitude response of the Dines instrument H_D was determined experimentally by Miller *et al.* (2013) and is shown in Figure 3.8. The typical response of the automatic anemometer H_{aws} is given, under certain assumptions (Miller *et al.*, 2013), by accounting for the filtering effect of using a finite time period $H_1(f)$, 3-second averaging $H_2(f)$, and the dynamic behaviour of the instrument itself $H_3(f)$ as

$$H_{aws}(f) = H_1(f)H_2(f)H_3(f), \tag{3.5}$$

where

$$H_1(f) = \left[1 - \left(\frac{\sin(T\pi f)}{T\pi f}\right)^2\right]^{1/2}, \tag{3.6}$$

$$H_2(f) = \left[\frac{|\sin(3\pi f)|}{3\pi f}\right], \tag{3.7}$$

and

$$H_3(f) = \left[1 + \left(\frac{2\pi f d}{\overline{v}}\right)^2\right]^{-1/2}. \tag{3.8}$$

Miller *et al.* (2013)'s correction factors were derived based on the mean Gust factor G (Davenport, 1964), defined as the ratio of peak gust to mean wind speed $\overline{v}$, given by

$$G = 1 + g\frac{\sigma_v}{\overline{v}}. \tag{3.9}$$

The g is a peak gust factor and σ_v is the standard deviation of the along-wind speed fluctuations about $\overline{v}$ over a time period T. If the along-wind speed fluctuations follow a Normal distribution, then Davenport (1964) showed

$$g = \sqrt{2\ln(\nu T)} + \frac{\gamma}{2\ln(\nu T)}, \tag{3.10}$$

where $\gamma = 0.577\ldots$ is Euler's constant. If the spectral density function of the wind speed fluctuations $S_v(f)$ is known, then σ_v and a cycling rate ν (Davenport, 1964) can be calculated as

$$\sigma_v^2 = \int_0^\infty S_v(f)df \tag{3.11}$$

and

$$\nu^2 = \frac{\int_0^\infty f^2 S_v(f)df}{\int_0^\infty S_v(f)df} \tag{3.12}$$

respectively. While any suitable expression for $S_v(f)$ could be used (Beljaars, 1987), Miller *et al.* (2013) used the von Kármán form

$$S_v(f) = \frac{4(\overline{v}I_v)^2(fl_v/\overline{v})}{f\left[1 + 70.8(fl_v/\overline{v})^2\right]^{5/6}}, \tag{3.13}$$

where l_v is an integral length scale, equal to 85m for a recording at 10m (Miller *et al.*, 2013).

Equation 3.13 requires the turbulence intensity I_v. This cannot be readily determined from the available data, but can be estimated from the terrain surrounding the station as

$$I_v = \frac{1}{\ln(z/z_o)}. \tag{3.14}$$

Where z is the instrument height and z_o is the surface roughness, as given in Section 3.3. Equation 3.14 performs poorly (Holmes, 2018) for thunderstorm winds. In the absence of more information, $I_v = 0.1$ was assumed for all thunderstorm gusts as recommended by Holmes and Ginger (2012).

Applications of Eqs. 3.9 $\rightarrow$ 3.14 allow a Gust factor to be calculated for a given $\overline{v}$ and z_o. The response of the anemometer can be factored into G by multiplying $S_v(f)$ by the square of an amplitude response function $H(f)$ so that

$$\sigma_{v^*}^2 = \int_0^\infty S_v(f)H(f)^2 df \tag{3.15}$$

and

$$\nu_*^2 = \frac{\int_0^\infty f^2 S_v(f)H(f)^2 df}{\int_0^\infty S_v(f)H(f)^2 df} \tag{3.16}$$

replace Eqs. 3.11 and 3.12 respectively.

The gust that would have been measured using the newer anemometer v_{aws} can then be estimated given a Dines measurement v_D as

$$v_{aws} = v_D \frac{G_{aws}}{G_D}. \tag{3.17}$$

3.4.2 Surface Roughness Estimation

To calculate the I_v and to correct for surface roughness, the z_o of the terrain around a station is required. To do this using the method in Section 3.3.2 requires a fairly precise location of a station. While precise locations could be determined post-1990, they could not be found for the period in which the Dines instruments were employed. The annual reports from the South African Weather Bureau (1950–1980) specifies the locations of stations in degrees and minutes. Therefore, the location of Dines instruments can only be determined to within roughly 1.5 km. Further uncertainty is introduced by inconsistency in locations over time, the South African Weather Bureau (1950–1980) indicate numerous minor (< 1.5 km) changes in instrument location.

To estimate the locations of the Dines instruments, the imprecise locations supplied

by the South African Weather Bureau (1950–1980) were used as a starting point. Google Earth's satellite imagery and historical aerial photography (Department of Rural Development and Land Reform, Republic of South Africa, 2020) were used to refine estimates of the instrumentation locations. The Dines instrumentation were housed in small buildings which were unlikely to be far from the weather office, because hourly recordings had to be made manually (Kruger, 2011).

The GIS approach, described in Section 3.3, could then be employed using the estimated locations and the 1990 land cover GIS layer (Department of Environmental Affairs Republic of South Africa, 2020). Determination of z_o for 30° segments, as applied in Section 3.3, was deemed inappropriate due to uncertainty in the exact location of the instrumentation. Instead 90° segments were employed, whose lower resolution should decrease reliance on precise locations and hopefully still yield acceptably representative estimates of z_o.

3.4.3 Verification of Correction

The method described in Section 3.4.1, using the surface roughness determined in 3.4.2, was applied to the Dines gust recordings at stations identified in Section 3.2.1. Uncertainty in the station locations and applicability to thunderstorm events meant that the accuracy of these corrections is uncertain.

To investigate the accuracy, the k-sample Anderson Darling test (Scholz and Stephens, 1987) was used. This tests the null hypothesis that multiple samples were drawn from the same population by comparing their empirical distribution functions. Testing if the corrected Dines data and AWS data at a given station could have originated from the same population provides an indication of the validity of a correction.

The empirical distribution functions and the data from the AWS, uncorrected Dines and corrected Dines stations are shown on Gumbel plots in Figure 3.B1, where P-values p_A resulting from the k-sample Anderson Darling test (Scholz and Stephens, 1987) are given in the corresponding legends. In all cases data was corrected for non-standard surface roughness and heights. The plots show that the strength of evidence against a

correction varies substantially. In some cases (Cape Town, Irene, Langebaan, Upington) there is insufficient evidence $p_A > 0.1$ to indicate an inadequate correction, while in others (Bloemfontein, Kimberly, Polokwane, Port Elizabeth) there is strong evidence $p_A < 0.05$ of inadequate correction.

It was decided to only include corrected Dines data in the expanded dataset that could not be rejected at a 5% significance level, i.e $p_A < 0.05$. These gusts were noted as originating from the Dines instrument so that if it is found that they lead to a significant difference in future results, the decision to include them can be re-evaluated. This should mitigate severe effects of an invalid correction.

The advantage of including the Dines Data is a substantial reduction in model-variance, which can be appreciated by the reduction in width of confidence intervals. This is demonstrated in Figure 3.9 where the Gumbel distribution was fit, using maximum likelihood estimation, to the AWS data alone and to the AWS data combined with the Dines data and then Confidence intervals were obtained using bootstrapping.

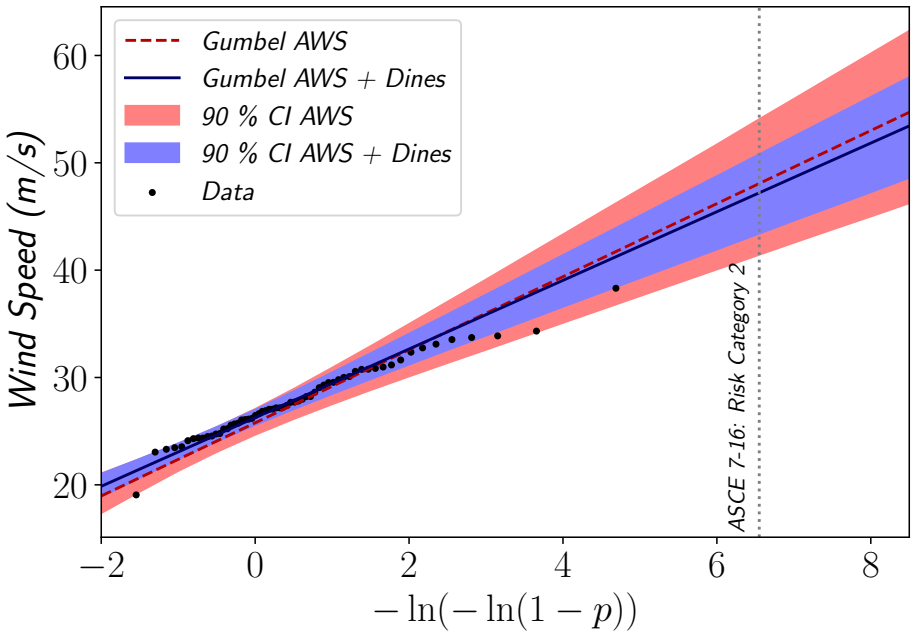


Figure 3.9: Impact of Dines data on Gumbel 90% confidence intervals at the Johannesburg Weather Office.

3.5 Conclusion

The result of this analysis is an expanded dataset containing around 3500 individual 3-second annual gust maxima, classified by climatic mechanism, which can be used for assessment of SANS 10160-3:2019. This dataset is built on the one created by Kruger (2011). Additional sources of data from the 10 years subsequent to Kruger (2011)'s study and from new stations have been included. Previously unused data measured using outdated Dines anemometers were also incorporated into the dataset. The proportion of each data source in the expanded dataset and the spatial distribution of stations is presented in Figures 3.10 and 3.11 respectively.

To handle the increased quantity of data, more efficient methods were investigated. In particular, the use of GIS to automate surface roughness correction. As large quantities of data will continuously become available, further automation of gust extraction from the SAWS database should be pursued. This will require automatic error correction and mechanism identification. The development of an automatic data extraction algorithm could use the roughly 3200 visually verified and classified gusts provided by this study as a training set.

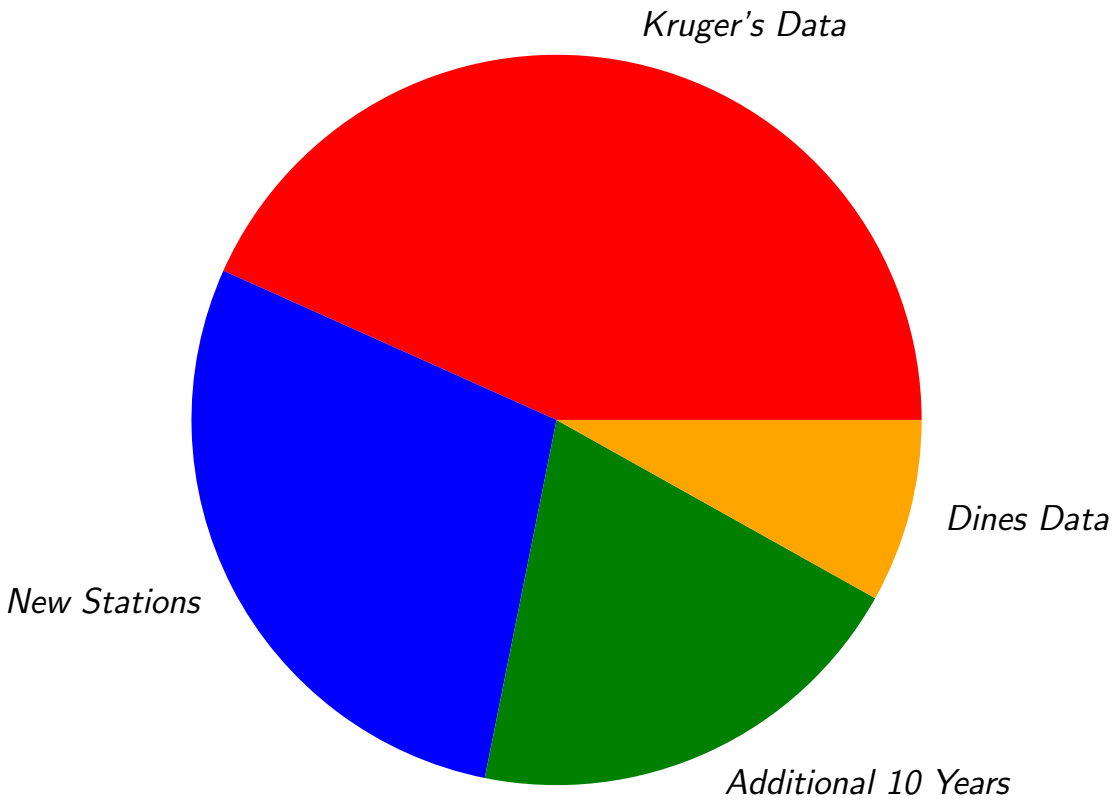
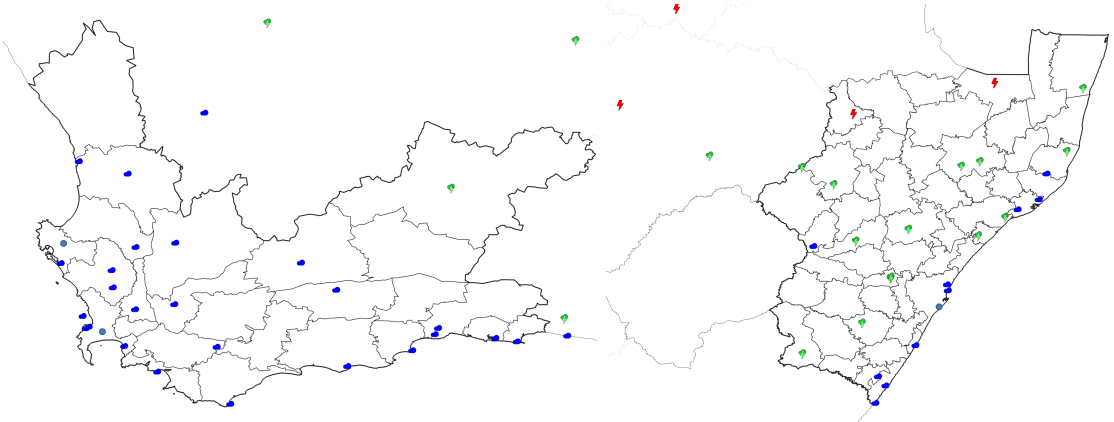
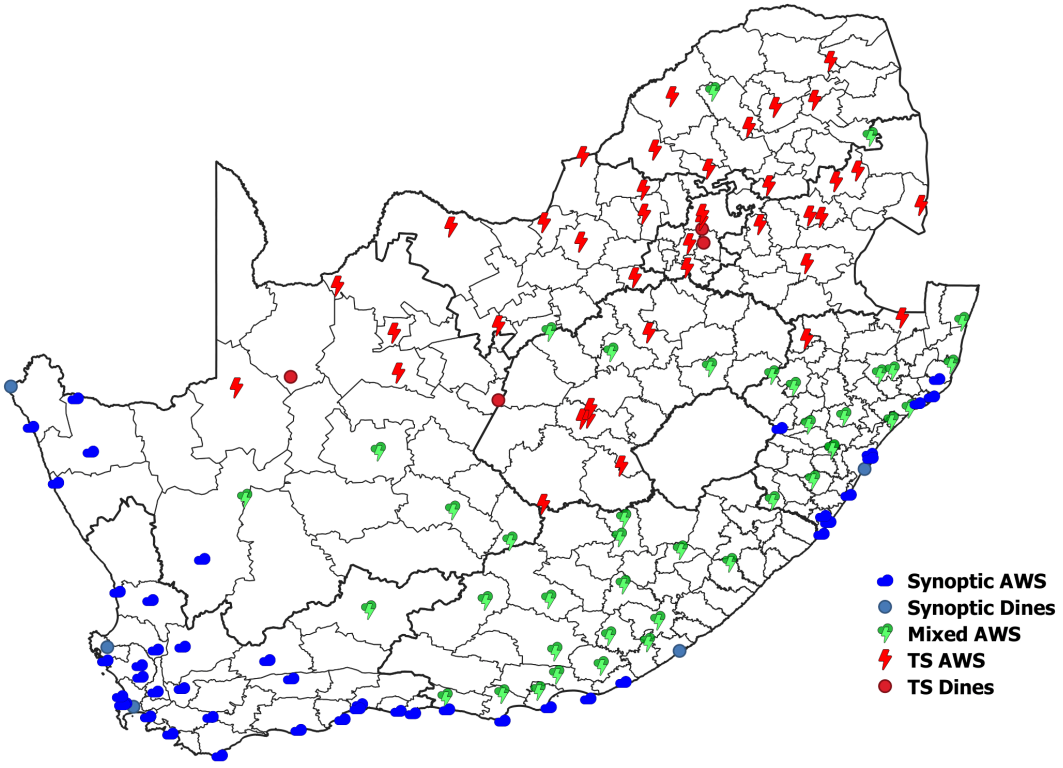


Figure 3.10: Sources of data in the new dataset.



(a) Western Cape Province (b) Kwa-Zulu Natal Province



(c) South Africa

Figure 3.11: Spatial distribution of stations in the new dataset, with thunderstorm (TS), mixed, synoptic, dines, and AWS indicated.

Appendix 3

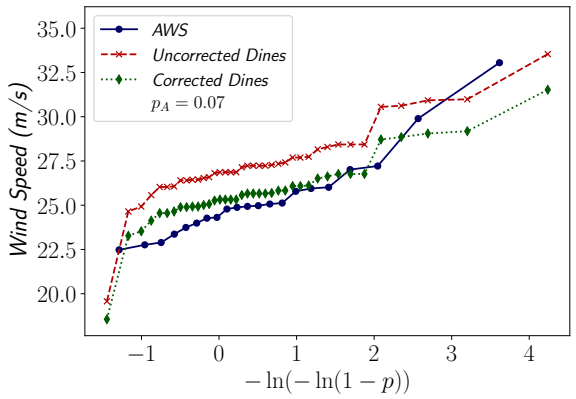
3.A Land Cover to Surface Roughness Conversion

No	Land Cover	z_o	Colour
1	Seasonal Water	0.005	
2	Permanent Water	0.0002	
3	Wetland	0.03	
4	Indigenous Forest	1	
5	Dense Bush, Thicket and Tall Dense Shrubs	0.5	
6	Woodland and Open Bushland	0.25	
7	Grassland	0.03	
8	Low Shrub land: Fynbos	0.1	
9	Low Shrub land: Other	0.03	
10	Commercial field high	0.25	
11	Commercial field medium	0.1	
12	Commercial field low	0.1	
13	Commercial pivot high	0.5	
14	Commercial pivot medium	0.1	
15	Commercial pivot low	0.1	
16	Orchards high	0.5	
17	Orchards medium	0.5	
18	Orchards low	0.25	
19	Vineyards high	0.25	
20	Vineyards medium	0.1	
21	Vineyards low	0.1	
22	Pineapples	0.03	
23	Subsistence Farm High	0.25	
24	Subsistence Farm medium	0.1	
25	Subsistence Farm low	0.1	
26	Sugar-cane pivot standing	0.1	
27	Sugar-cane pivot fallow	0.1	
28	Sugar-cane non-pivot standing	0.1	
29	Sugar-cane non-pivot fallow	0.1	
30	Sugar-cane emerging standing	0.1	
31	Sugar-cane emerging fallow	0.1	
32	Forest Plantations: Mature Trees	1	

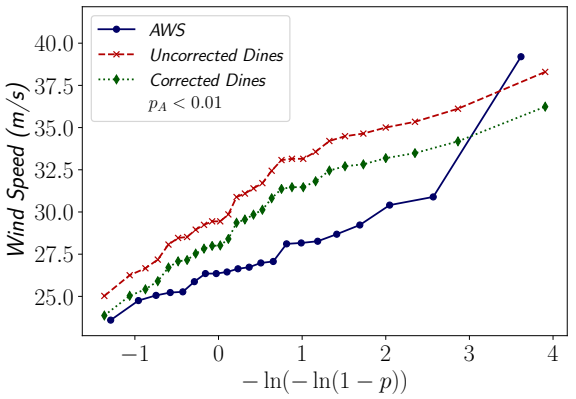
33	Forest Plantations: Young Trees	0.5	
34	Forest Plantations: Clear-felled	0.5	
35	Mine Bare	0.03	
36	Mine Semi-bare	0.03	
37	Mine water seasonal	0.005	
38	Mine water permanent	0.0002	
39	Mine Buildings	1	
40	Erosion Dongas and Gullies	0.005	
41	Bare (Non Vegetated)	0.005	
42	Built Up: Commercial	2	
43	Built Up Industrial	2	
44	Built Up Informal: tree dominated	1	
45	Built Up Informal: bush dominated	1	
46	Built Up Informal: grass dominated	0.5	
47	Built Up Informal: bare dominated	0.5	
48	Built Up Residential: tree dominated	1	
49	Built Up Residential: bush dominated	1	
50	Built Up Residential: grass dominated	0.5	
51	Built Up Residential: bare dominated	0.5	
52	Schools and Sports Grounds	0.5	
53	Built Up Smallholding: tree dominated	0.5	
54	Built Up Smallholding: bush dominated	0.25	
55	Built Up Smallholding: grass dominated	0.25	
56	Built Up Smallholding: bare dominated	0.25	
57	Built Up Sport and Golf: tree dominated	0.5	
58	Built Up Sport and Golf: bush dominated	0.25	
59	Built Up Sport and Golf: grass dominated	0.1	
60	Built Up Sport and Golf: bare dominated	0.03	
61	Built Up Township: tree dominated	1	
62	Built Up Township: bush dominated	1	
63	Built Up Township: grass dominated	0.5	
64	Built Up Township: bare dominated	0.5	
65	Built Up Village: tree dominated	1	
66	Built Up Village: bush dominated	0.5	
67	Built Up Village: grass dominated	0.1	
68	Built Up Village: bare dominated	0.1	
69	Built Up: tree dominated	1	
70	Built Up: bush dominated	0.5	
71	Built Up: grass dominated	0.25	
72	Built Up: bare dominated	0.1	

Table 3.A1: Land cover classes available from Department of Environmental Affairs Republic of South Africa (2020), with assigned roughness length z_o.

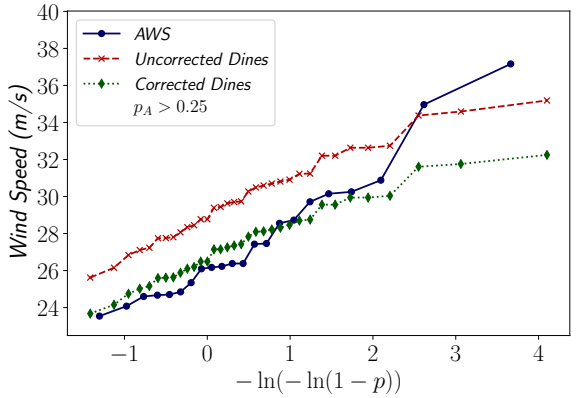
3.B Empirical CDF with Anderson Darling Results of Dines Corrections



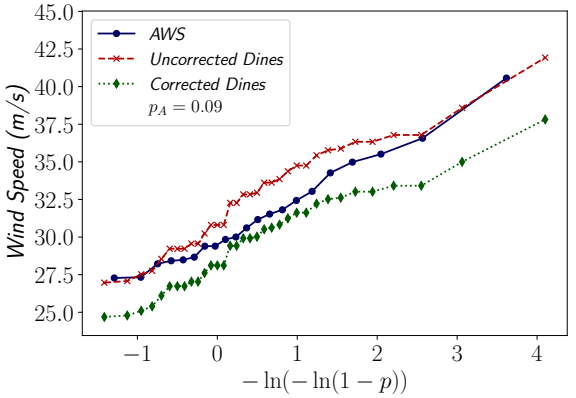
(a) Alexanderbaai



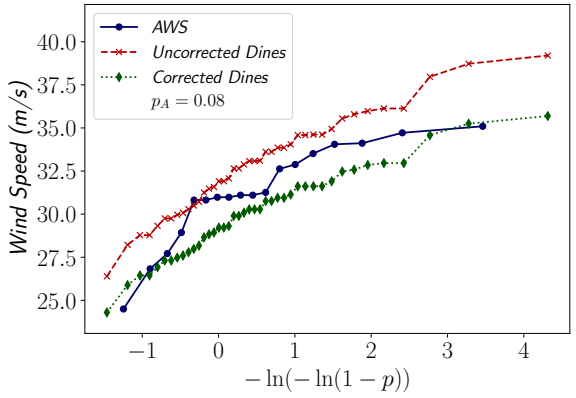
(b) Bloemfontein



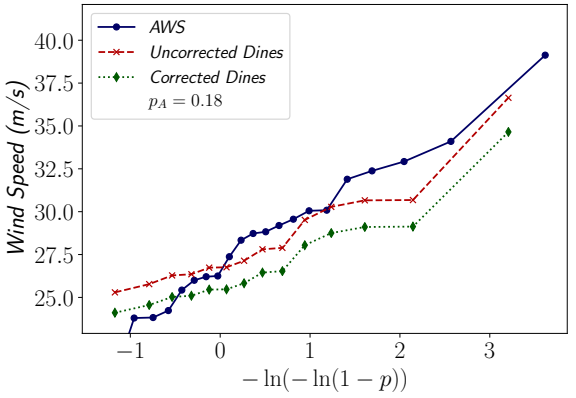
(c) Cape Town



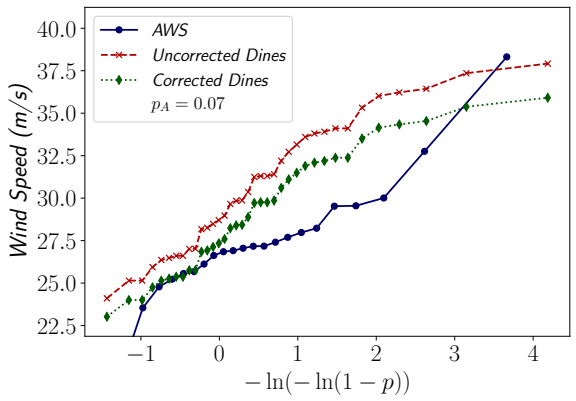
(d) Durban



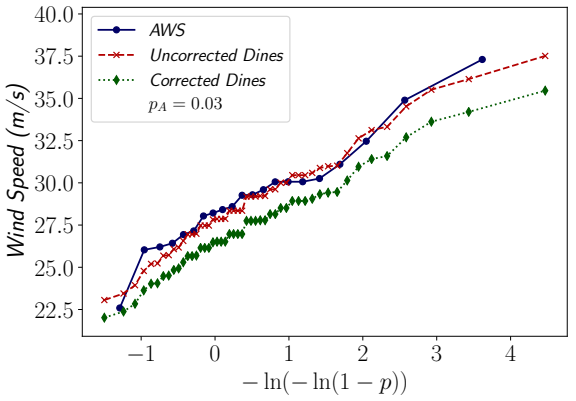
(e) East London



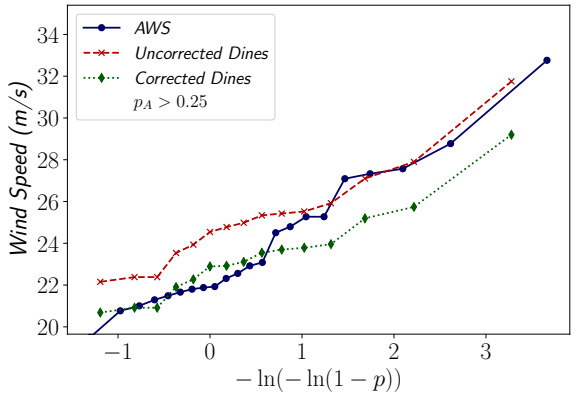
(f) Irene (Pretoria)



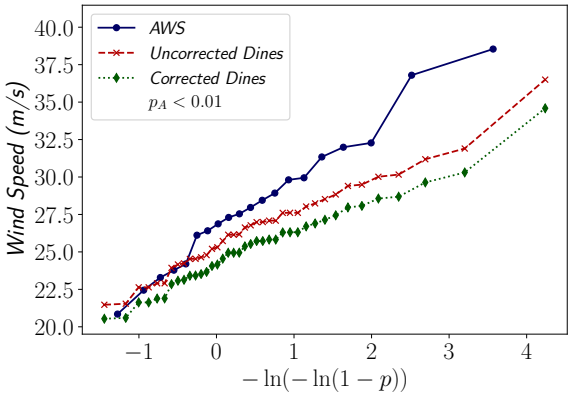
(g) Johannesburg



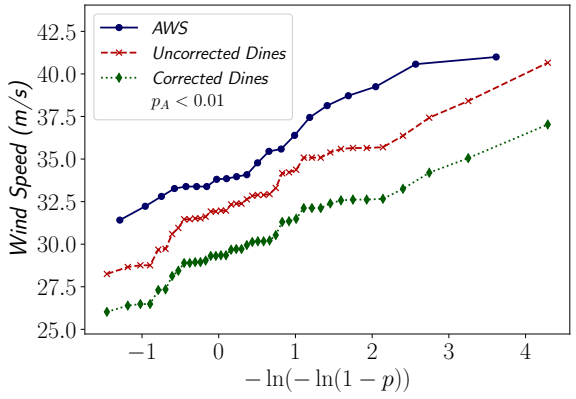
(h) Kimberly



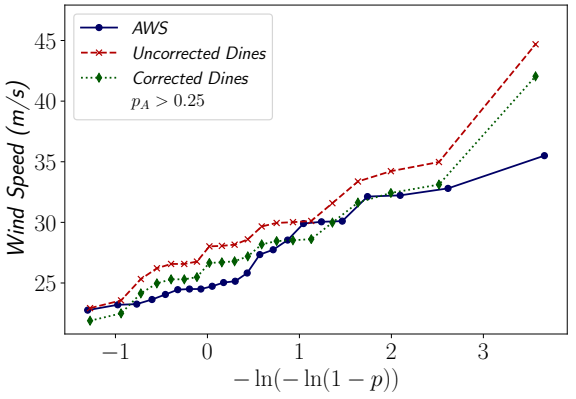
(i) Langebaan



(j) Polokwane (Pietersburg)



(k) Port Elizabeth



(l) Upington

Figure 3.B1: Empirical distribution functions and the data from the AWS, uncorrected Dines and corrected Dines data at selected stations. The P-value resulting from the k-sample Anderson Darling test (Scholz and Stephens, 1987) is given in the corresponding legend, for each case.

Chapter 4

Regional Distribution Selection

FP Bakker, N de Koker, C Viljoen

Submitted to *Engineering Structures* as *Preconditioning Wind Speeds for Standardised Structural Design.*

Summary

The statistical treatment of historical wind data has a significant impact on the design wind load, standardisation of which requires characterisation of the wind climate across a large region. Improved spatial resolution is afforded by asymptotic extreme value distributions as they require a lower temporal data resolution. The Gumbel distribution is shown to be preferred, given the available South African data, according to the Akaike Information Criterion. However, its relative inflexibility means that estimates of small probability fractiles could be significantly biased, a difficulty illustrated by disagreement on which is more appropriate: fitting the Gumbel distribution to the wind speed versus to the wind-induced pressure (squared wind speed). By generalising this choice to the problem of finding the most appropriate real-valued exponent to raise the wind speeds by (i.e. not just 1 or 2), this paper aims to reduce this model-bias. As estimates of this pre-conditioning parameter from a single station would have low confidence given the typical quantity of data, a method of estimating the exponent that maximises the likelihood of

observing the entire dataset is developed instead. This regional preconditioning method is demonstrated using over 3500 annual gust measurements at 131 stations throughout South Africa, the most likely exponent is found to be 1.59 with confidence intervals sufficiently narrow to reject fitting to both the wind speed as well as to the squared wind speed. It is shown that using an exponent of 1.6 in deriving the design wind load for structural design would result in a substantial reduction in model-bias, providing a suitable baseline for future South African loading standards.

Notation

q_d	Design wind pressure
v_k	Characteristic wind speed
γ_w	Wind load partial factor
c_k	Wind load modification factor
p	Exceedance probability
v	Wind speed
x	Generic random variable
u_x	GEV location parameter of x
b_x	GEV scale parameter of x
k_x	GEV shape parameter of x
n	Sample size
h	Number of parameters
l	Log likelihood
$AICc$	Small sample adjusted Akaike Information Criterion
$q_d \vert (v \sim Gumbel)$	Design wind pressure given Gumbel distributed wind speeds
v_d	Design wind speed
v^w	Preconditioned wind speed
w	Preconditioning parameter/Weibull shape parameter
$\overline{w}$	Regional preconditioning parameter
N	Number of series in dataset
$\mathbf{V}$	Wind speed dataset with N series
$\vec{V_i}$	Series of wind speeds
$\vec{Z_i}$	Preconditioned series of wind speeds
$\hat{w}$	Estimated preconditioning parameter
$\hat{\overline{w}}$	Estimated regional preconditioning parameter
$\mathbf{V_{SA}}$	South African dataset
$\mathbf{V_{ts}}$	Thunderstorm subset of South African dataset
$\mathbf{V_{syn}}$	Synoptic subset of South African dataset
$\overline{w}'$	Assumed preconditioning parameter
s	Weibull scale parameter
z	Random variable after a transformation
γ	Euler's constant (0.57721...)

4.1 Introduction

4.1.1 Codified Wind Loading

The wind induced pressure q_d used for structural design is directly proportional to the square of the wind speed. It is generally represented in standards by

$$q_d = \gamma_w c_k v_k^2, \tag{4.1}$$

where the characteristic wind speed v_k is determined from observed speeds and is adjusted to an appropriate design value using the partial factor γ_w. The factor c_k represents various other components of the Davenport (1982) wind load chain, including the change in speed due to height, topography, air density and aerodynamics of the structure.

The codification of wind loads has been described in more detail in Section 2.4 and 2.5. An overview is repeated below.

Designers normally obtain v_k from a map or table in a national standard. In drawing up such standards, v_k is calculated from measured wind data for a set probability of exceedance p_k, assuming an appropriate distribution function. The v_k values are derived for a specific region using wind recordings in that region. These recordings are normalised by removing the effect of local topography and surface roughness. This allows the magnitude of the design wind load to be tailored to the site conditions using the factor c_k.

γ_w is the partial safety factor used to adjust the characteristic wind load to a design value with an appropriate level of reliability. γ_w is found by first assuming probability distributions, based on typical statistics, for the components in the wind loading process. The γ_w can then be determined by ensuring that q_d achieves a target level of reliability using a reliability analysis (Botha *et al.*, 2018*a*; Holickỳ, 2009; Ang and Tang, 1984).

One option to reduce model-bias (Burnham and Anderson, 2002) explored in recent work (Cook *et al.*, 2011; Hong *et al.*, 2016) as an alternative to the above formulation, involves a site-based design approach in which a high return period design wind speed

v_d is specified such that the characteristic value corresponds with the target reliability, effectively setting the partial factor to one.

In either formulation the design load q_d is dependant on the choice of distribution used to relate an exceedance probability p to the wind speed v, required to infer v_k and γ_w or v_d. Many such distributions have been suggested in the literature (Palutikof *et al.*, 1999; Holmes, 2018). It is important to note that an appropriate distribution choice will be dependant on the data that is available.

4.1.2 South African Dataset

The origin of the available dataset is discussed in more details in Chapter 3. An overview is repeated below.

In South Africa, wind speed data is collected and stored by the South African Weather Service (SAWS) using RM Young anemometers at over 200 stations. Kruger (2011) extracted a dataset of annual maximum 3-second gusts from this network to derive v_k for the South African wind loading standard (SANS 10160-3:2019). An issue that has been identified in Kruger (2011)'s work is short series lengths, with the longest being only 16 years. Cook (1985) recommends at least 20 years for reliable results. Building on the data used by Kruger (2011), an updated dataset of extreme gusts was extracted from the SAWS database in Chapter 3. This was done with the aim of further assessing SANS 10160-3 using extended series lengths and better spatial resolution. The use of gust data is required to capture thunderstorm events (Kruger, 2011; Milford, 1987*b*).

Kruger (2011) found that both thunderstorms and extratropical cyclones dominate the extreme wind climate of South Africa. This means that extreme wind speed data needs to be segregated by mechanism to allow for the proper statistical treatment recommended by Gomes and Vickery (1978). Thunderstorm gusts were identified by finding extremes that exhibited the following characteristics: a high ratio of gust to 5-minute mean wind speeds, the commencement of rain or hail, a temperature decrease, a shift in the wind direction and a pressure and humidity increase (Kruger, 2011).

While segregating data from the SAWS database both Kruger (2011) and Chapter 3

found spurious spikes in the wind record, these were not accompanied by any expected corresponding climatic information from pressure, humidity or rainfall sensors. This can be explained as obvious cases of instrumentation malfunction (Kruger, 2011). A large proportion of recorded strong wind gusts were deemed erroneous and excluded from the dataset, thus demonstrating that verification of the SAWS data is necessary.

To ensure sufficient geographical coverage, it was aimed to extract data from as many stations as possible. Given the labour required in the separation and verification process it was deemed most efficient to focus only on extraction of annual maxima. Cook (1985) recommend that sample sizes less than 10 years are of limited value and so stations with a record length less than 10 years were not considered.

The resulting dataset includes around 3500 annual maxima from 131[1] stations, shown in Figure 4.1. It was found that around 30% of stations contained both synoptic and thunderstorm maxima. For these stations two annual maximum data series were extracted; one synoptic and one thunderstorm which meant that the 131 stations yielded 169 series of annual maximum wind speeds. The wind speeds were then normalised by correcting for non-standard surface roughness following World Meteorological Organization (2008) recommendations, where surface roughness was determined by adapting a visual inspection to a Geographic Information System approach.

[1]In Chapter 3, 132 stations are mentioned. Originally the data from the Hoedspruit station was not transferred from Kruger (2011)'s dataset to the expanded dataset. This error has been corrected and was found to have no impact on the results of the analysis detailed in this Chapter.

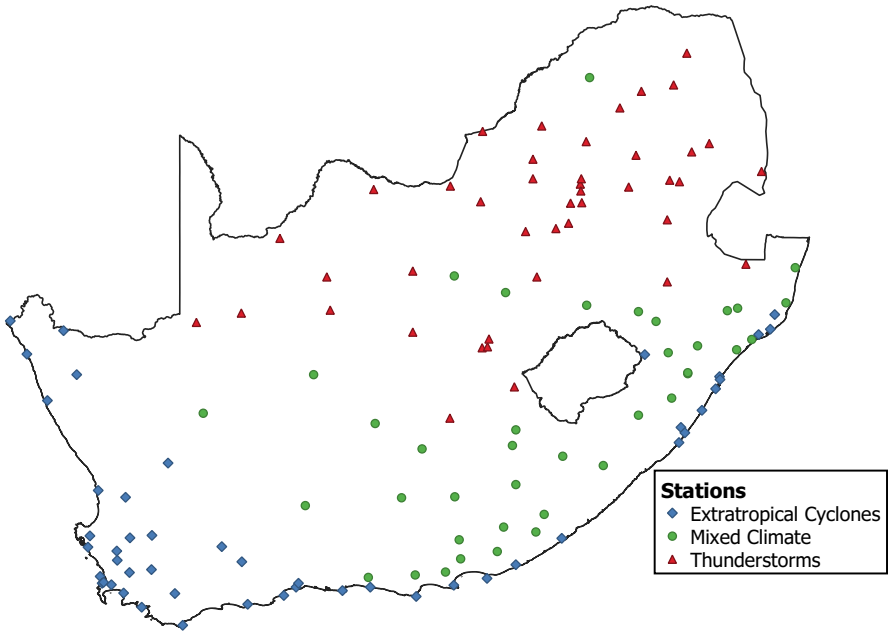


Figure 4.1: Distribution and dominant climatic mechanism of stations used in this study.

4.2 Statistical Representation of Extreme Wind

Codification requires selection of an appropriate probability distribution which adequately represents the extreme wind climate. Maximising the observations per site using a Peaks-Over-Threshold or Method of Independent Storms approach is likely to result in a more stable, and possibly more accurate, representation of the extreme wind behaviour at that site (Palutikof *et al.*, 1999). However, a practical limitation is that data from hundreds of stations throughout a region may need to be quality controlled and normalised, which becomes less feasible as the rate that data is collected continues to increase.

Asymptotic extreme value distributions require a limited number of extremes and so decreases the burden of quality control per station. This may allow data from more stations to be considered and yield better spatial resolution. Therefore, it is argued that to more accurately capture the extreme wind climate of a large region, where quality control of data is a significant constraint, restricting the analysis to annual extremes and asymptotic extreme value theory is reasonable. A similar argument is made by Hong *et al.* (2014) when reviewing climatic input data for the Canadian standard.

4.2.1 Asymptotic Extreme Value Distributions and Model Selection

Fisher and Tippett (1928) showed that, given a sufficient number of events, the largest can be described by one of three asymptotic types. These asymptotic types were combined into one form by Von Mises (1936) as the Generalised Extreme Value (GEV) distribution, with cumulative distribution function of

$$F_x(x) = \begin{cases} \exp\left[-\exp\left(-\frac{x-u_x}{b_x}\right)\right] & k_x = 0 \\ \exp\left[-\left(1-k_x\frac{x-u_x}{b_x}\right)^{\frac{1}{k_x}}\right] & k_x \neq 0. \end{cases} \tag{4.2}$$

The GEV distribution has a location, scale and shape parameter denoted by u_x, b_x and k_x, respectively. k_x controls the asymptotic behaviour with $k_x = 0$ indicating the Gumbel distribution and $k_x > 0$ the Reverse Weibull distribution.

The Gumbel distribution fixes the shape parameter and so is a two parameter distribution; while the Reverse Weibull requires three parameters and has an upper bound on x. The increased flexibility afforded by the additional parameter should generally result in a less biased model than the Gumbel, although at the cost of additional model-variance (Burnham and Anderson, 2002).

For a dataset of size n, the trade-off between the number of distribution parameters h and extent to which a model fits the data (measured by maximum log likelihood l) can be quantified by the small sample ($n/h_{max} < 40$) adjusted Akaike Information Criterion $AICc$ (2.3.5.1) (Akaike, 1973; Hurvich and Tsai, 1989)

$$AICc = -2l + \frac{2hn}{n-h-1}. \tag{4.3}$$

Note therefore that when comparing competing candidate models, the one which yields the smallest $AICc$ would be preferred.

Given this dataset, an indication of the probability of model m being the best of M

possible models is given by the the Akaike Weight (Burnham and Anderson, 2002)

$$\text{Akaike Weight} = \frac{\exp[(AICc_{min} - AICc_m)/2]}{\sum_{i=1}^{M} \exp[(AICc_{min} - AICc_i)/2]}. \tag{4.4}$$

This provides a measure of the strength of the evidence for a particular candidate model m relative to the M models under consideration.

Note that $n/h_{max} < 40$ for all the series in the South African dataset when comparing the Gumbel to either of the other asymptotic extreme value distributions.

4.2.2 Appropriate Extreme Value Distribution for Wind Data

Annual extreme winds have typically been modelled using the Reverse Weibull or Gumbel distributions (Holmes, 2018). The indication of the Fréchet form ($k_x < 0$) is not considered to be physically representative and is explained as the result of multiple unsegregated mechanisms (Palutikof *et al.*, 1999; Kruger, 2011; Gomes and Vickery, 1978). The use of a Reverse Weibull distribution for annual wind speed maxima generally results in the best fit (Kruger, 2011; Simiu *et al.*, 2001; Walshaw, 1994; Lechner *et al.*, 1992). It is further justified by reasoning that there must be some upper limit that the atmosphere could produce, which corresponds to the Reverse Weibull asymptotic behaviour (Holmes, 2018), although the prescription of an upper bound could have dangerous design implications when derived using limited data.

The simplicity and versatility of the Gumbel distribution has meant that it has seen widespread application in codification efforts (Botha *et al.*, 2018*a*; Kruger, 2011; Hong *et al.*, 2014; JCSS Model Code), it is unlimited and so the dangers of prescription of an upper bound are avoided. The typical indication of the Reverse Weibull form means that the Gumbel distribution is often conservative especially at high return periods (Holmes, 2018; Rózsás and Sỳkora, 2016).

Using the $AICc$ to compare the Gumbel versus the Reverse Weibull distribution, given the South African data, yields a lower $AICc$ for 85% of the 169 series. The Akaike Weights, shown in Figure 4.2, indicate a clear preference for the Gumbel distribution.

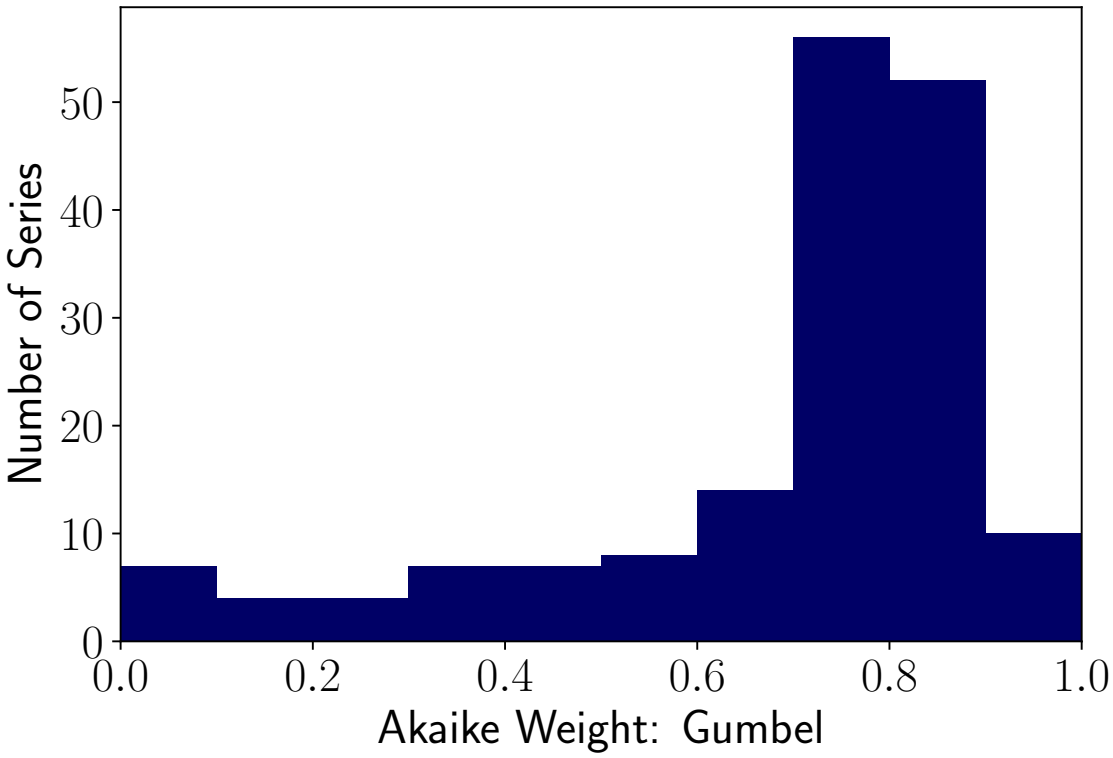
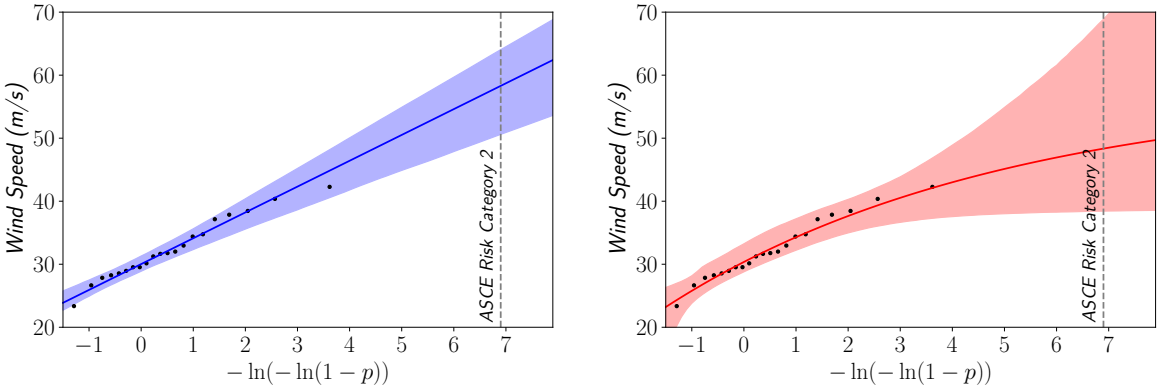


Figure 4.2: Akaike Weight of the Gumbel vs the Reverse Weibull Distribution for all 169 series in the South African dataset.

The GEV is equivalent to the Reverse Weibull distribution with the additional flexibility of $k_x < 0$, which is considered physically unrepresentative (Palutikof *et al.*, 1999; Kruger, 2011; Gomes and Vickery, 1978). Therefore, the Gumbel distribution is also preferred over the GEV distribution in South Africa. The same conclusions were reached by Hong *et al.* (2014) when reviewing appropriate distributions for updating the Canadian standard.

These results can be interpreted in the context of the bias-variance trade off (Friedman *et al.*, 2001). The additional flexibility afforded by the shape parameter of the GEV distribution means that it should generally fit better than the Gumbel (lower model-bias) but at the cost of additional model-variance. The $AICc$ can be used to quantify this trade-off and indicates that on average lower prediction error should be realised using the Gumbel distribution, despite the additional model-bias associated with fixing the shape parameter.



(a) Gumbel (b) GEV

Figure 4.3: The 90% confidence intervals when the Gumbel and GEV distribution are applied to annual maxima from the Strand weather station. The confidence intervals were found using parametric bootstrapping.

Measures of model performance like $AICc$ correspond to the general predicative ability of a model (Burnham and Anderson, 2002). The intention when fitting extreme wind speed values for design is to make predictions in the tail of the distribution, where the uncertainty is higher. This means that the $AICc$ is likely to under-penalise additional parameters and that the practical performance (extrapolation capability) of the Gumbel distribution relative to the GEV distribution is likely to be even better than indicated by the $AICc$ (Laio *et al.*, 2009). The unsuitability of the GEV distribution for extrapolation is demonstrated in Figure 4.3 where the width of the GEV confidence interval is significantly greater than the Gumbel confidence interval at higher return periods.

4.2.3 Application of the Gumbel Distribution: Wind Speed or Squared Wind Speed?

When considering the Gumbel distribution as a viable option for updating SANS 10160-3, it was unclear whether it should be applied to the wind speed or the dynamic wind pressure. The current v_k in SANS 10160-3:2019 was found by fitting the Gumbel distribution to wind speeds at 76 stations throughout South Africa (Kruger, 2011). Botha *et al.* (2018*a*) then performed a reliability assessment of SANS 10160-3 to determine γ_w, where the wind pressure was assumed to follow a Gumbel distribution.

The relationship (Eq. 4.1) between wind pressure and v imply statistical inference of v^2 is equivalent to inference of wind pressure, if uncertainties in c_k are ignored. Assuming that v or v^2 is Gumbel distributed leads to different results, with fitting to v giving notably more conservative estimates (Hong *et al.*, 2014; Botha, 2016), as demonstrated in Figure 4.5.

Fitting a Gumbel distribution to v^2 was suggested by Cook (1982) and justified by the shape parameter of the Forwards Weibull distributed parent. Milford (1987*a*) supported fitting to v^2, based on agreement with estimates derived from the parent distribution throughout South Africa. Conversely fitting the Gumbel distribution to v^2 was assessed by Simiu *et al.* (2001), who concluded that there is insufficient evidence for extreme dynamic pressures following the Gumbel distribution. In a discussion on whether to fit to v or v^2 for the derivation of Canadian wind loads, Hong *et al.* (2014) decided to fit to v. This was justified based on inspection of probability plots and the tendency to err on the conservative side. The question of the appropriate statistical representation of Gumbel distributed extreme wind loading remains unresolved.

When assessing probabilistic models, Botha (2016) found that most assumed the wind pressure followed a Gumbel distribution (Holickỳ, 2009; JCSS Model Code; Gulvanessian and Holicky, 2005; Milford, 1985*a*), although not all did so (Ellingwood and Tekie, 1999). To be consistent with other models, Botha *et al.* (2018*a*) used a Gumbel distribution to represent wind pressure when performing a reliability calibration of the South African code. Given the uncertainty about the correct approach, the impact of fitting to v or v^2 in the context of design wind loading is investigated.

The impact of fitting the Gumbel distribution to v or v^2 can be assessed by the resulting difference in the design value. A measure of this is given by

$$\Delta_{(1-2)} = \frac{q_d|(v \sim Gumbel) - q_d|(v^2 \sim Gumbel)}{q_d|(v^2 \sim Gumbel)}, \tag{4.5}$$

where $q_d|(v \sim Gumbel)$ is the design load when v is assumed to be Gumbel distributed and $q_d|(v^2 \sim Gumbel)$ is the design load when v^2 is assumed to be Gumbel distributed.

The q_d is dependant on a partial factor γ_w and a characteristic wind speed v_k (Eq. 4.1). Some standards (AS/NZS 1170.2; ISO 4354:2009; ASCE 7-16) specify a high return period design wind speed v_d that corresponds with the target reliability, effectively setting the partial factor to one. In this case the design wind speed is given by $q_d = c_k v_d^2$ and

$$\Delta_{(1-2)} = \frac{v_d^2|(v \sim Gumbel) - v_d^2|(v^2 \sim Gumbel)}{v_d^2|(v^2 \sim Gumbel)}. \tag{4.6}$$

In ASCE 7-16, v_d is defined as the 700-year return period wind speed for risk category II structures (which cover the majority of structures) (McAllister *et al.*, 2018). If this return period is accepted as indicative of typical design requirements, then $\Delta_{(1-2)}$ can be calculated for each of the 131 available stations. This was done by fitting the Gumbel distribution to v and to v^2 using Lieblein (1976)'s Best Linear Unbiased Estimators technique (BLUE) and accounting for mixed climates, as recommended by Gomes and Vickery (1978), when necessary. These fits were used to extrapolate to a 700-year return period to find v_d, which could then be used to find $\Delta_{(1-2)}$ with Eq. 4.6. The distribution of $\Delta_{(1-2)}$ from all the stations is shown in Figure 4.4.

On average $\Delta_{(1-2)}$ was found to be 12 % but could be as high as 35 %. This effect would differ dependant on the risk category in ASCE 7-16, as shown in Figure 4.5, and would increase for risk category III and IV structures, associated with 1700 and 3000 year return periods respectively. While South African design requirements SANS 10160-3:2019 are not equivalent to ASCE 7-16, the results suggest that, regardless of the precise requirements, the assumption of whether to fit to v or v^2 is significant enough to substantially impact design values.

4.2.4 Preconditioning Wind Speeds

The primary issue with using the Gumbel distribution is that the inflexibility of a fixed shape means that substantial model-bias could be introduced, especially for high return periods. This issue is compounded by uncertainty in whether to fit to v or v^2 since the incorrect choice could substantially increase model-bias.

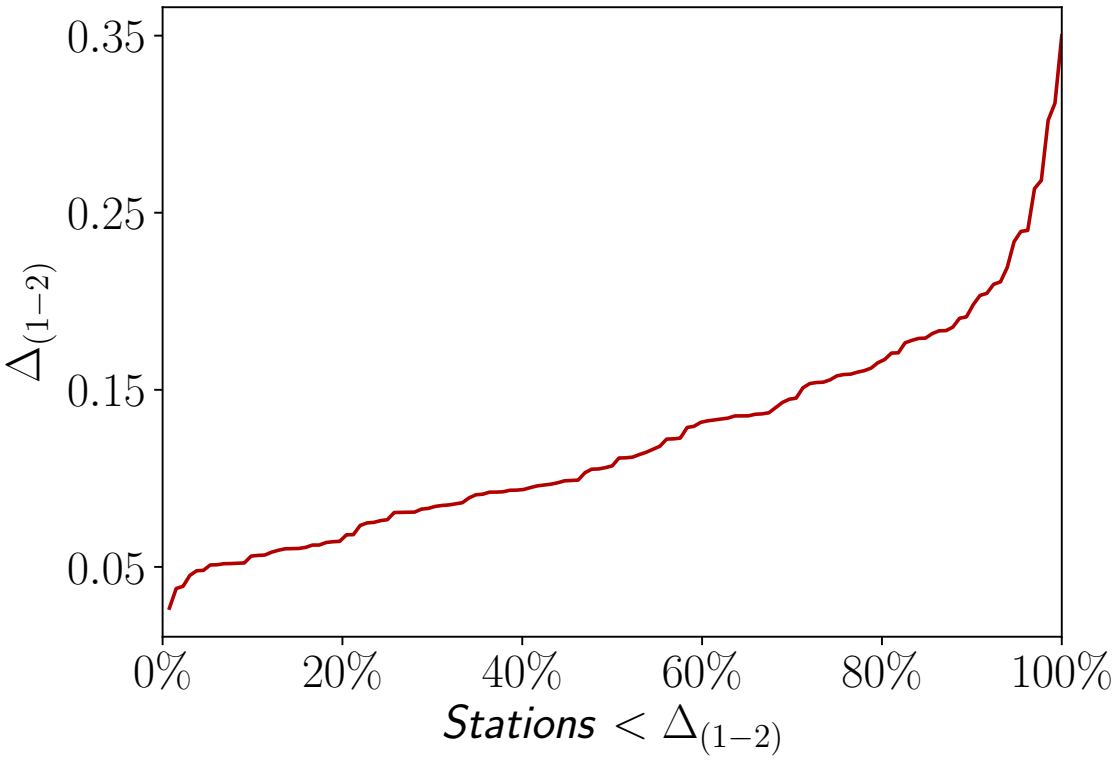


Figure 4.4: Difference between fitting to v and v^2 for available stations, measured by $\Delta_{(1-2)}$.

The Gumbel distribution is the appropriate asymptotic distribution for extremes of events whose parent has an exponential tail (Castillo *et al.*, 2005). If this was true for v it would still be true for v^2 although fitting to v leads to different results than v^2. This is because the convergence to the Gumbel is affected by the transformation of v to v^2. It has been suggested that fitting the Gumbel to v raised to a power w equal to the Weibull shape parameter fitted to the parent would improve convergence (Cook and Harris, 2004), and therefore yield more accurate fractile estimates. This argument is given in more detail in Appendix 4.A.

Obtaining estimates of v by fitting the Gumbel distribution to v^w to improve convergence is termed *preconditioning*, which Cook and Harris (2004) show is equivalent to a General Penultimate Gumbel distribution with a cumulative distribution function given by

$$F_x(x) = \exp\left[-\exp\left(-\frac{x^w - u^w}{b^w}\right)\right]. \tag{4.7}$$

On a Gumbel plot of v, preconditioning can be visualised as adding curvature to the plot. This means w is effectively a shape parameter and is analogous to k_x from the GEV distribution, with $w = 1 \implies k_x = 0$, $w < 1 \implies k_x < 0$ and $w > 1 \implies k_x > 0$, (Holmes, 2018). The similarity between the General Penultimate (Eq. 4.7) and GEV (Eq. 4.2) distributions is demonstrated in Figure 4.5, where fits were obtained using maximum likelihood estimation on the Johannesburg Weather Office data. The dominance of the Reverse Weibull form ($k_x > 0$), relative to wind speed, means in general it is expected for $w > 1$. An advantage of preconditioning as a substitute for the Reverse Weibull is that the prescription of an upper bound is avoided, so that as one extrapolates to higher return periods the tail is thicker, maintaining a tendency to conservatism.

If w is treated as a distribution parameter and estimated from the data, then the deviations from $w = 1$ and $w = 2$ could provide insight into the applicability of fitting the Gumbel distribution to v and v^2, respectively.

Two approaches of obtaining w are considered by Cook and Harris (2004). The first is to estimate it directly from the annual extremes, this is equivalent, although parametrisation differs, to the procedure investigated by Hong (2015) to obtain consistent estimates regardless of fitting to v or v^2.

This method was tested by individually fitting the General Penultimate distribution (Eq. 4.7), using maximum likelihood estimation, to each of the 169 available series. The $AICc$ scores were then calculated, which indicate that performance is similar to the GEV distribution. The Gumbel distribution was preferred (gave a lower $AICc$) for 91 % of the stations. Therefore, directly estimating w from the extremes appears to offer little advantage over the GEV or Reverse Weibull, because like estimating k given a limited series length, estimating w introduces substantial model-variance.

The second approach considered by Cook and Harris (2004) is to first fit the Forwards Weibull distribution to the parent to find w before preconditioning. Estimates of w using the parent distribution are based on far more data and so the issue of increased model-variance is effectively avoided. Defining the parent distribution of the extreme winds is difficult since they are comparatively rare events and so need to be segregated

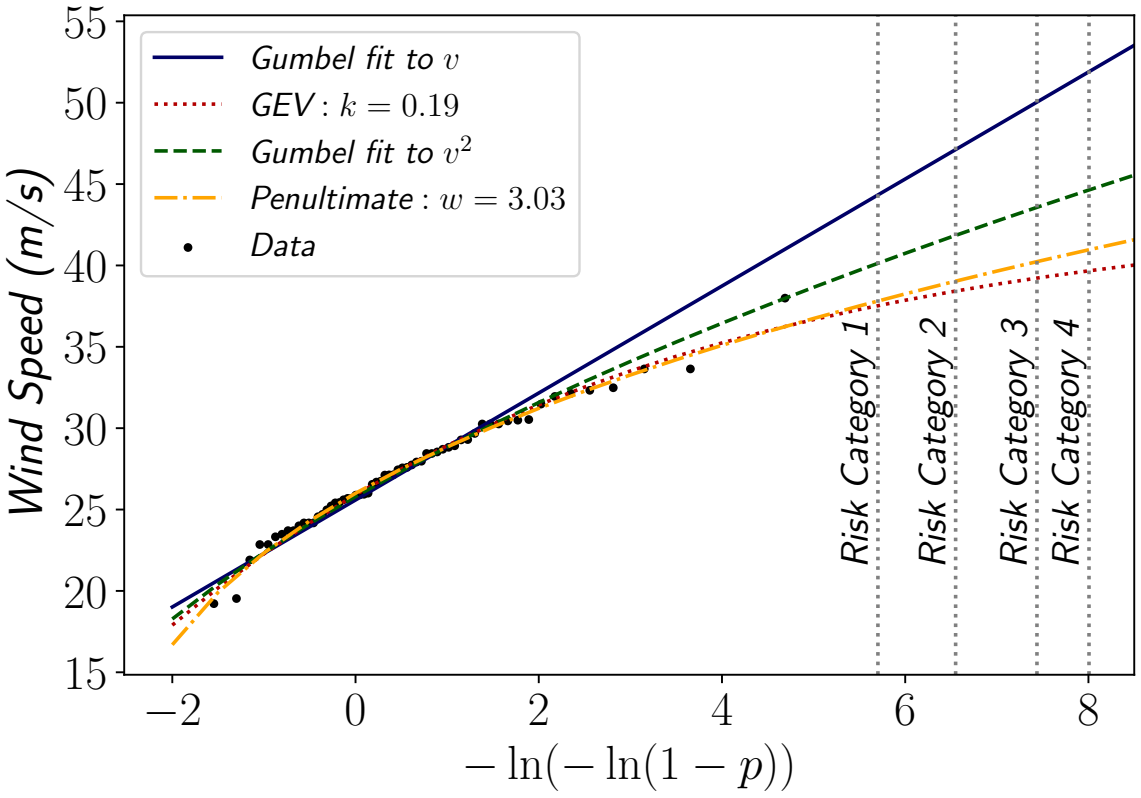


Figure 4.5: Distributions fitted to data collected at the Johannesburg Weather Office, shown on Gumbel Paper. Distributions were fit using maximum likelihood estimation and data is plotted using Gringorten (1963)'s plotting positions.

from the recordings at each station (Holmes, 2018; Cook and Harris, 2008). Attempting to accurately separate the parent at the hundreds of available stations in the SAWS database was deemed unfeasible.

According to Holmes (2018) the typical range of w values found in the parent is between 1.3 and 2 and values within this range were generally encountered in the literature (Cook, 1982; Milford, 1987*a*; Ayodele *et al.*, 2012; Cook *et al.*, 2003; Torrielli *et al.*, 2013). When w was fit to only a single series using the General Penultimate distribution for the Johannesburg station $w = 3.03$ was obtained (Figure 4.5). Values of w this high are not typically encountered in the parent distribution. This result is likely due to the model-variance associated with using only a single series of annual maximums.

4.3 Estimation of a Regional Preconditioning Parameter $\overline{w}$

Since determining w at each station, as recommended by Cook and Harris (2004), is not considered practical, fixing w at a generally appropriate value $\overline{w}$ for a region is investigated. Estimation of this regional $\overline{w}$ could use data from multiple stations thus decreasing the model-variance compared with site by site estimation. Application of an appropriate $\overline{w}$ for a region would result in decreased model-bias compared to not preconditioning, where it is arbitrarily assumed that $\overline{w} = 1$ ($v \sim$ Gumbel) or $\overline{w} = 2$ ($v^2 \sim$ Gumbel).

A similar regional estimation of the GEV shape parameter k_x could be developed and is expected to achieve similar results to regional preconditioning. In this study only regional preconditioning is pursued, because of its stronger theoretical justification based on the typical Forwards Weibull parent distribution, and because it does not result in an upper bound and is thus more conservative than the GEV for high return period extrapolation.

4.3.1 Estimation Procedure

A procedure which incorporates data from multiple series to estimate a regional preconditioning parameter $\overline{w}$, while accounting for the series by series estimation of the Gumbel parameters u_z and b_z of the preconditioned series $z = v^{\overline{w}}$, was developed. This procedure is based on maximum likelihood estimation, which aims to find the parameters for which the observations have the highest likelihood of occurrence.

A wind speed data set $\mathbf{V}$ is made up of N series of recordings $\{\vec{V_1}, \vec{V_2}, ...\vec{V_N}\}$ with each individual series $\vec{V_i}$ having n_i observations $\{v_1, v_2....v_{n_i}\}$. Preconditioning the data by $\overline{w}$ results in $\{\vec{Z_1}, \vec{Z_i}....\vec{Z_N}\}$ where each $\vec{Z_i}$ is $\{v_1^{\overline{w}}, v_2^{\overline{w}}....v_{n_i}^{\overline{w}}\}$.

It is assumed that any preconditioned series $\vec{Z_i}$ can be described by a Gumbel distribution. Therefore the probability density function of the random variable z is

$$f_z(z) = \frac{1}{b_z} \exp\left[\frac{u_z - z}{b_z} - \exp\left(\frac{u_z - z}{b_z}\right)\right], \tag{4.8}$$

where estimates of u_z and b_z are determined using a Gumbel parameter estimation procedure on $\vec{Z}_i$ (Hong *et al.*, 2013). This accounts for the separate series by series estimation of the Gumbel parameters.

Another issue is that the maximum likelihood estimation must be performed on a common scale. To ensure that likelihood is relative to v, a transformation of variables (Benaroya *et al.*, 2005) is used to relate the probability density function of v to $f_z(z)$ as

$$\begin{aligned} f_v(v|\overline{w}) &= \frac{dz}{dv} f_z(z) \\ &= \frac{\overline{w} v^{\overline{w}-1}}{b_z} \exp\left[\frac{u_z - v^{\overline{w}}}{b_z} - \exp\left(\frac{u_z - v^{\overline{w}}}{b_z}\right)\right]. \end{aligned} \tag{4.9}$$

If it is assumed that the observations are independent, then the log likelihood of $\overline{w}$ can be calculated

$$l(\overline{w}|\mathbf{V}) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} \ln[f(v_j|\overline{w})], \tag{4.10}$$

and an estimate of $\overline{w}$ obtained

$$\hat{\overline{w}} = \underset{\overline{w}}{\operatorname{argmax}}[l(\overline{w}|\mathbf{V})]. \tag{4.11}$$

The maximum can be found using a suitable numerical optimization algorithm (Nelder and Mead, 1965; Snyman, 2005). An advantage of using maximum likelihood estimation is that a $100(1-\alpha)\%$ confidence interval can be found using the likelihood ratio by solving for $\overline{w}$ in

$$2[l(\overline{w}|\mathbf{V}) - l(\hat{\overline{w}}|\mathbf{V})] = \chi^2_{1-\alpha}, \tag{4.12}$$

where $\chi^2_{1-\alpha}$ is the $(1-\alpha)$ quantile of a chi square distribution with a single degree of freedom.

4.3.2 Application to South African Data

To demonstrate the regional preconditioning estimation procedure developed above, it was applied to the full South African dataset $\mathbf{V_{SA}}$, where u_z and b_z were determined

using separate maximum likelihood estimation per series. The region in this case is South Africa. It was found that $\hat{\overline{w}}(\mathbf{V_{SA}}) = 1.59$ with $CI_{90\%}\{1.40, 1.79\}$, the likelihood function is shown in Figure 4.6. This result is well within the bounds of typical values of w observed in wind speed parent distributions (Holmes, 2018; Cook, 1982; Milford, 1987*a*; Ayodele *et al.*, 2012; Cook *et al.*, 2003; Torrielli *et al.*, 2013). The confidence intervals indicate that both the hypotheses $\hat{\overline{w}} = 1$ and $\hat{\overline{w}} = 2$ can be rejected with high confidence ($P - \text{value} << 0.01$).

To ensure that the estimation procedure is acceptably robust, different estimation procedures of u_z and b_z were tested. Using Lieblein (1976)'s BLUE technique, the same $\hat{\overline{w}}(\mathbf{V_{SA}})$ result was obtained as maximum likelihood, demonstrating that the $\hat{\overline{w}}$ obtained is fairly insensitive to the selected fitting procedure for u_z and b_z.

Cook and Harris (2004) point out the possibility of geographical variation of $\overline{w}$ in the United States. The dominance of synoptic events and thunderstorms in different regions of South Africa, as shown in Figure 4.1, mean that the same $\overline{w}$ may not be appropriate for the entire country. To test this hypothesis $\mathbf{V_{SA}}$ was split into a thunderstorm subset $\mathbf{V_{ts}}$ of 82 series and a synoptic subset $\mathbf{V_{syn}}$ of 87 series. The $\overline{w}$ was then estimated for each and found to be $\hat{\overline{w}}(\mathbf{V_{ts}}) = 1.53$ with $CI_{90\%}\{1.27, 1.79\}$ and $\hat{\overline{w}}(\mathbf{V_{syn}}) = 1.68$ with $CI_{90\%}\{1.38, 1.99\}$. While $\hat{\overline{w}}(\mathbf{V_{ts}}) < \hat{\overline{w}}(\mathbf{V_{syn}})$, there is low confidence in the outcome.

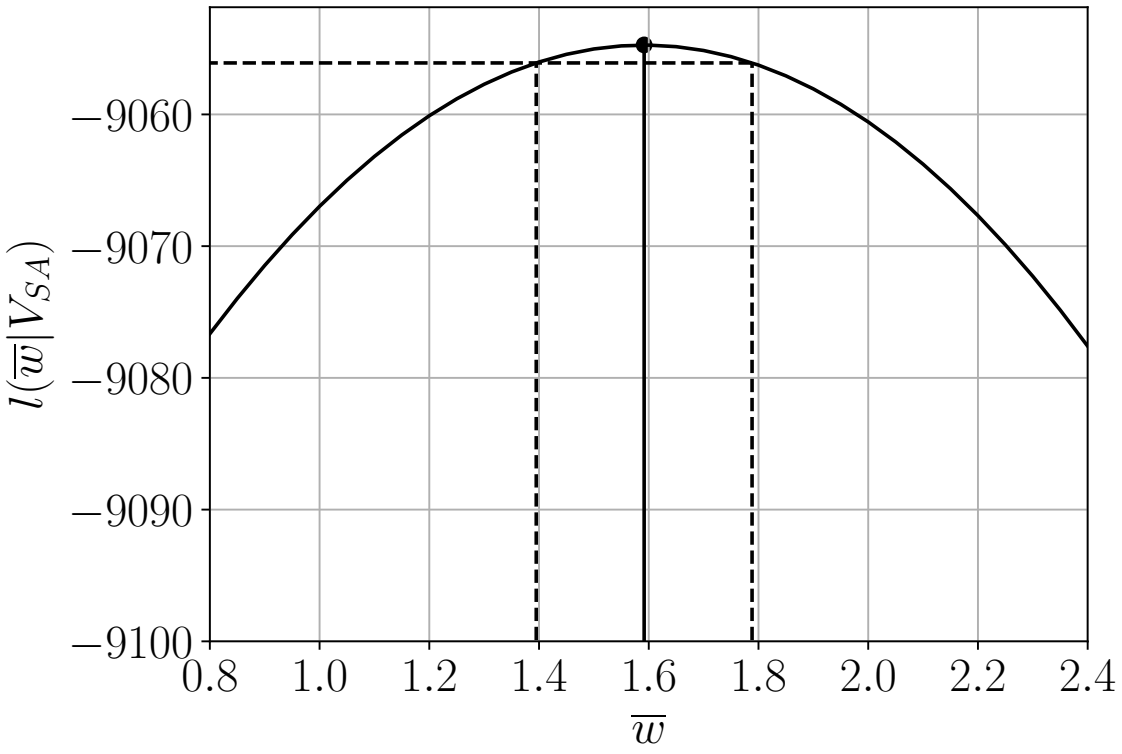


Figure 4.6: Log-likelihood of $\overline{w}$ for $\mathbf{V}_{SA}$.

4.4 Model Bias when Regionally Preconditioning

The impact of regional preconditioning can be informed using the insights into likely values of $\overline{w}$ and assessed through the design wind pressure q_d. To do this Eq. 4.6 can be generalised as

$$\Delta_{(\overline{w}'-\overline{w})} = \frac{v_d^2|(v^{\overline{w}'} \sim Gumbel) - v_d^2|(v^{\overline{w}} \sim Gumbel)}{v_d^2|(v^{\overline{w}} \sim Gumbel)}. \tag{4.13}$$

where $v_d|(v^{\overline{w}} \sim Gumbel)$ is the appropriate design wind speed, obtained by fitting a Gumbel distribution to $v^{\overline{w}}$. $v_d|(v^{\overline{w}'} \sim Gumbel)$ represents the design wind speed that would be obtained when some other $\overline{w}$ is assumed $\overline{w}'$ and a Gumbel distribution fit to $v^{\overline{w}'}$ instead.

The $\hat{\overline{w}}$ from the maximum likelihood procedure can be used to estimate the model-bias when $\overline{w}$ is assumed to be $\overline{w}'$. Adopting the ASCE 7-16 700-year return period for v_d,

the average model-bias in fitting to $v, v^{1.6}$ and v^2, given the South African data, can be obtained by setting $\overline{w} = \hat{\overline{w}}$ and $\overline{w'} = \{1, 1.6, 2\}$ in Eq. 4.13.

The results, presented in Figure 4.7, show that fitting a Gumbel distribution to v should result in greater model-bias than fitting to v^2. The average $\Delta_{(\overline{w}'-\overline{w})}$ was between a 5 and 10 % overestimation when fitting to v and between a 2 and 6 % underestimation when fitting to v^2. These errors increased substantially at some stations, the confidence intervals indicate that the model-bias could be as high as a 29 % overestimation or a 14 % underestimation when fitting to v or v^2 respectively. These results would be more pronounced for stricter reliability requirements.

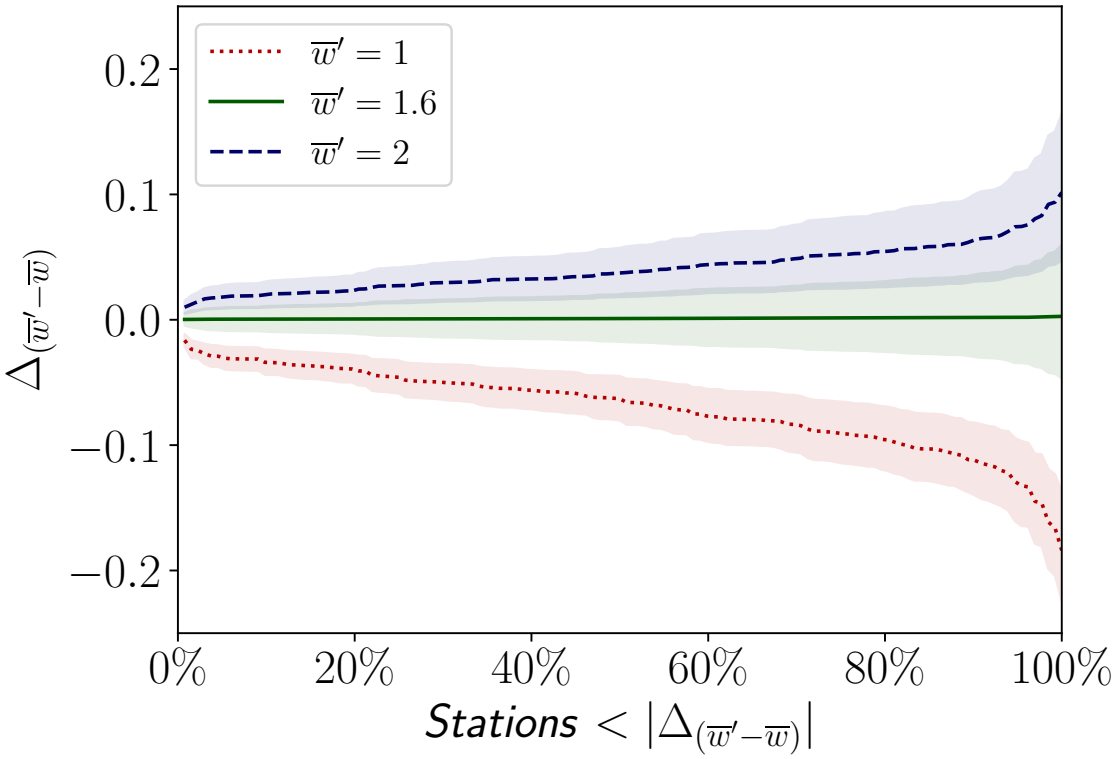


Figure 4.7: model-bias when fitting to $v^{\overline{w}'}$.

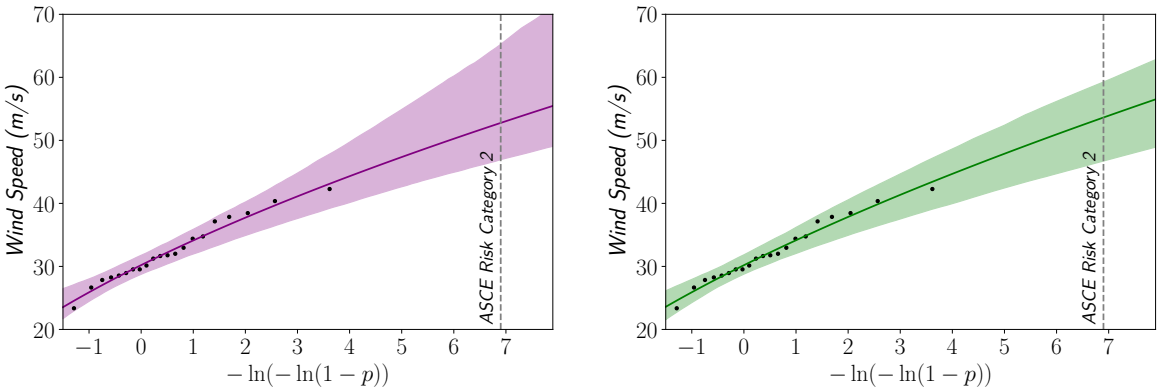
Fitting to v^2 should result in less model-bias than fitting to v, but the empirical result of $\hat{\overline{w}} = 1.59$ indicates that doing so is likely to result in under-design. Fitting to $v^{1.6}$ could be considered as a fairly unbiased option. Figure 4.7 shows that fitting to $v^{1.6}$ results in substantially lower model-bias than either of the other options, with an average $\Delta_{(\overline{w}'-\overline{w})}$

within 2 % and a $\Delta_{(\overline{w}'-\overline{w})}$ less than 6 % in an extreme case. This exercise also shows that the impact of the uncertainty in $\hat{\overline{w}}$ is minimal, since within the 90% confidence interval it is likely to impact the design value by less that 2%.

4.5 Model Variance when Regionally Preconditioning

The additional flexibility afforded by the preconditioning exponent reduces the model-bias associated with the use of the Gumbel distribution. If this exponent were estimated using only the data from a single series this could significantly increase the model-variance. This is shown in Figure 4.8a where the confidence intervals suggest a similar level of uncertainty as the GEV distribution (Figure 4.3b) when only a single series is used for estimation.

Incorporating more information by estimating a single $\overline{w}$ for a number of series should reduce the additional model-variance introduced by the estimation. If the number of series used to estimate $\overline{w}$ is large enough, then this additional model-variance may be insignificant. When $\overline{w}$ was estimated using all the available South African annual maximum wind speed series, an estimate of $\hat{\overline{w}} = 1.59$ was obtained. The uncertainty of this estimate



(a) Single series preconditioned Gumbel.

(b) Regionally preconditioned Gumbel.

Figure 4.8: The 90% confidence intervals when the preconditioned Gumbel distribution is applied to Annual Maxima from Stand Weather station. The confidence intervals were found using parametric bootstrapping.

was quantified using the log-likelihood function, which indicated a standard error of 0.13. This is substantially less than when a single series is used: at the Strand weather station when the w was estimated, the log-likelihood function indicates a standard error of 2.0.

The uncertainty when regionally preconditioning can be investigated using parametric bootstrapping. This is shown in Figure 4.8b for the Strand weather station. Comparing this to Figure 4.8a, where a w was estimated using only the annual maxima of the Strand station, the width of the 90% confidence intervals are substantially reduced. The width of the confidence intervals in Figure 4.8a suggests a similar level of uncertainty to using the Gumbel distribution (Figure 4.3a) and a significant reduction compared to the GEV distribution (Figure 4.3b). The curvature introduced by preconditioning $\overline{w} = 1.6$ improves the fit compared to the Gumbel distribution (Figure 4.3a) by better following the typical Reverse Weibull form.

4.6 Conclusion

The purpose of a design code is to define a model that can be used by practising engineers to capture physical reality. To facilitate a meaningful conservative approach to design, this model should have a baseline that is as unbiased as possible. Appropriate design values consistent with the uncertainties of the model can then be assigned.

The Gumbel distribution is simple and generally applicable, which makes it an attractive option for codification purposes. Using the AIC_c measure it was shown to be preferred over the GEV distribution, given the available data in South Africa. The fact that a shape parameter does not need to be estimated reduces the model-variance substantially when applied to limited series lengths. The downside of this is that this inflexibility introduces model-bias, an issue which is compounded by uncertainty in whether to apply the Gumbel distribution to the wind speeds or to the wind induced pressure.

Preconditioning is the transformation of the wind speed v with some exponent w and then fitting the Gumbel distribution to v^w and can be justified theoretically based on the parent distribution. Preconditioning can be compared to the GEV distribution with

the preconditioning parameter acting as an effective substitute for the shape parameter, although unlike the GEV distribution preconditioning avoids the prescription of an upper bound.

To avoid estimation of the shape parameter for each series and the corresponding increase in model-variance, a single regional preconditioning parameter $\overline{w}$ could be estimated for the entire data set. This would reduce the overall model-bias compared to arbitrarily fitting to the wind speed or squared wind speed.

A procedure based on maximum likelihood estimation was developed to explore, from the annual maximums, which regional preconditioning exponent $\overline{w}$ is most appropriate. It was found that the most likely $\overline{w}$ was 1.59 given the over 3500 annual maxima in the South African dataset. This estimate falls comfortably within the typical w range found in the parent distributions.

The impact on design, given the results of the maximum likelihood estimation, for an assumption of $\overline{w}$ was assessed. It was found that fitting to the wind speeds preconditioned by $\overline{w} = 1.6$ is likely to significantly reduce model-bias compared to fitting to the wind speed or wind pressure in South Africa. Regionally preconditioning the wind speeds was found to lead to no significant increase in model-variance when the entire South African dataset was used for estimation.

The maximum likelihood method demonstrated in this paper could be applied to other national wind data sets to provide an insight as to which $\overline{w}$ values can be justified empirically if considering the Gumbel distribution.

Appendix 4

4.A Relationship between Preconditioning and the Parent Distribution

The Gumbel distribution is most appropriate for the extremes from a block of an exponentially distributed parent distribution (Cook and Harris, 2004). The Gumbel distribution is valid for extremes from parent distributions with an exponential tail (Castillo *et al.*, 2005). Many commonly used distributions have an exponential tail, including the Normal and Forwards Weibull, although the rate of convergence to the Gumbel varies. As a result, an important caveat is that the extreme must be from a sufficient number of independent events.

The parent distribution of wind speeds are generally best described by a Forwards Weibull distribution (Holmes, 2018; Cook and Harris, 2004; Harris and Cook, 2014)

$$F(v) = 1 - \exp\left[-\left(\frac{v}{s}\right)^{w}\right], \tag{4.A1}$$

where s is the scale and w is the shape parameter. This justifies the use of the Gumbel distribution for the extreme, since the Forwards Weibull distribution has an exponential tail. Despite this, the use of the Gumbel distribution for annual extreme wind speeds has been shown to potentially introduce increasing conservative model-bias at higher return periods (Walshaw, 1994; Lechner *et al.*, 1992; Rózsás and Sỳkora, 2016; Cook, 1982).

Cook and Harris (2004) consider the Forwards Weibull parent distribution and advo-

cate for the substitution $z = v^w$, then

$$F(z) = 1 - \exp\left[-\frac{z}{s_z}\right] \tag{4.A2}$$

and z is exponentially distributed. Hence the extremes of z converge rapidly resulting in a better fit of the Gumbel distribution. This fit could then be used to obtain a fractile of v from the corresponding fractile of z. This procedure is termed *preconditioning*.

Chapter 5

Optimal Variability Estimation

FP Bakker, N de Koker, C Viljoen

Submitted to *Structural Safety* as *Site or Regional Design Wind Speeds?*

Summary

Design values that achieve an acceptable probability of failure can be found by fitting an appropriate distribution to observed data. For the design wind load this is complicated by spatial variation of the wind climate, therefore it is unclear which data are relevant at a particular location. We aim to frame this problem in terms of model selection and the bias variance trade-off. Existing models for incorporating site versus regionally averaged statistics are expressed in terms of key parameters, which allow the errors associated with each model to be derived. It is found that the use of a characteristic wind speed in codification combines the site statistics (low bias model) with regionally averaged statistics (low variance model), with the return period of the characteristic wind speed shown to act as a parameter controlling the relative weighting of these models. This insight is used to develop an optimal (minimum mean square error) estimator of the design wind load, which varies based on both the available quantity of data at a particular site and how well this data corresponds to the regional average. The practical implications of each modelling approach are demonstrated at several South African stations.

Notation

r	Resistance of a structure
g	Permanent loads on a structure
q	Variable loads on a structure
v	Wind speed
c	Wind load modification factors
ρ	Air density
v_d	Design wind speed
p	Probability of exceedance
p_t	Target probability of exceedance
ε^2	Mean Squared Error (MSE)
ε^2_{tot}	Total MSE
ε^2_{bias}	Portion of MSE attributed to model-bias
ε^2_{std}	Portion of MSE attributed to model-variance
ε^2_{irr}	Irreducible portion of MSE
u_v	Gumbel location parameter of annual maximum wind speeds
b_v	Gumbel scale parameter of annual maximum wind speeds
μ_v	Mean of annual maximum wind speeds
σ_v	Standard deviation of annual maximum wind speeds
θ	Standardisation function for a Gumbel distribution
θ_t	Number of standard deviations away from the mean of the design value
n	Number of annual maxima in a sample from a single site
$\hat{\mu}_v$	Estimated mean
$\hat{\sigma}_v$	Estimated standard deviation
$\hat{v}_{d:s}$	The v_d estimated using site statistics
$\hat{v}_{d:r}$	The v_d estimated using regional average coefficient of variation
$\sqrt{\gamma_w}'$	Regional frequency factor
$\sqrt{\gamma_w}$	Square root of wind load partial safety factor neglecting uncertainty in c
δ_v	Site coefficient of variation
$\overline{\delta}_v$	Regional coefficient of variation
N	Number of sites in the region
$\hat{\delta}_v$	Estimated coefficient of variation from a sample
p_k	Annual probability of exceedance of the characteristic value, typically 0.02
v_k	Characteristic wind speed
t_k	Return period of the characteristic value, typically 50-years
$\hat{v}_{d:k}$	The v_d estimated using a Characteristic Design
θ_k	Standard score of the characteristic value, typically 2.59
$\hat{v}_{d:a}$	Design wind speed estimated using any modelling approach
ε_a	Normalised Root MSE (NRMSE) associated with a particular model
λ_v	Skewness of annual maximum wind speeds
κ_v	Kurtosis of annual maximum wind speeds
ε_s	NRMSE associated with a Site Design
ε_r	NRMSE associated with a Regional Design
ε_k	NRMSE associated with a Characteristic Design
$\theta_{k:opt}$	Standard score of the estimated optimal characteristic value
$\hat{v}_{d:o}$	Design wind speed estimated using the optimal characteristic value
$\hat{\theta}_{k:opt}$	Standard score of the optimal characteristic value
ε_o	NRMSE associated with an Optimised Design
q_d	Design wind pressure
γ	Euler's constant (0.57721...)

5.1 Introduction

5.1.1 Codified Wind Loading

Standardisation in structural design aims to ensure that the resistance provided by a structure exceeds the loading that the structure can be reasonably expected to be exposed to. This requirement can be expressed as

$$r > g + q, \tag{5.1}$$

where r, g and q are random variables representing relevant components of a structural system. The r and g are the resistance and permanent loads of the structure, and q represents the variable loads. The wind pressure is one such variable load and is related to wind speed v as

$$q = \frac{1}{2}\rho c v^2, \tag{5.2}$$

where ρ is the air density and c represents various components of the Davenport wind load chain, such as the change in pressure due to height, topography and aerodynamics of the structure (Davenport, 1982).

The probability that the resistance exceeds the loading, i.e. that Eq. 5.1 does not hold, is termed the probability of failure. A reliability-based wind loading standard attempts to map a specified target probability of failure to a set of optimal design values. Deviation from this target probability should be minimised throughout the scope of the standard (Retief and Dunaiski, 2009).

The probability distribution of r, g and c can be considered location independent while v is location dependant (Ellingwood and Tekie, 1999), which has made finding design values for the wind speed v_d more complicated, and is the reason for the use of wind speed maps in standards (SANS 10160-3:2019; ASCE 7-16; AS/NZS 1170.2; EN 1991-1-4).

If some tolerance for deviation from the design point is permitted (Cook *et al.*, 2011), then the target reliability of the wind load, i.e. annual probability of exceedance p_t, can be

separated from other components (Section 5.A). This approach has been adopted by the American ASCE 7-16 standard and has been shown to result in more uniform attainment of the target reliability (Hong *et al.*, 2016). The task of specifying appropriate design values using this approach requires a model to predict v_d for a given p_t throughout the area covered by the standard.

5.1.2 Modelling and Error

The aim of a model is to adequately describe the relationship between two quantities so that predictions, as near to the actual value as possible, can be made. The quality of a model can thus be assessed by the mean squared difference between a prediction and the actual value (MSE), this implies a valid approach to model selection would be to find the model that minimises MSE.

The total MSE ε^2_{tot} can be expressed in term of three types of error (Friedman *et al.*, 2001)

$$\varepsilon^2_{tot} = \varepsilon^2_{bias} + \varepsilon^2_{std} + \varepsilon^2_{irr}. \tag{5.3}$$

The ε_{bias} is the average difference between the model prediction and the actual value (model-bias), the ε^2_{std} is the variance of the predictions (model-variance) and ε^2_{irr} represents irreducible errors outside the scope of available models. Since all models should be equally affected by ε^2_{irr} it can be neglected for the purposes of model selection.

There is generally a relationship between model complexity and MSE, with model-bias increasing and model-variance decreasing with model complexity. Further it has been found that there typically exists a trade-off between the model-bias and model-variance, with the best model balancing the model-bias and model-variance by utilising an optimal level of complexity (Friedman *et al.*, 2001). This concept is demonstrated in Figure 5.1.

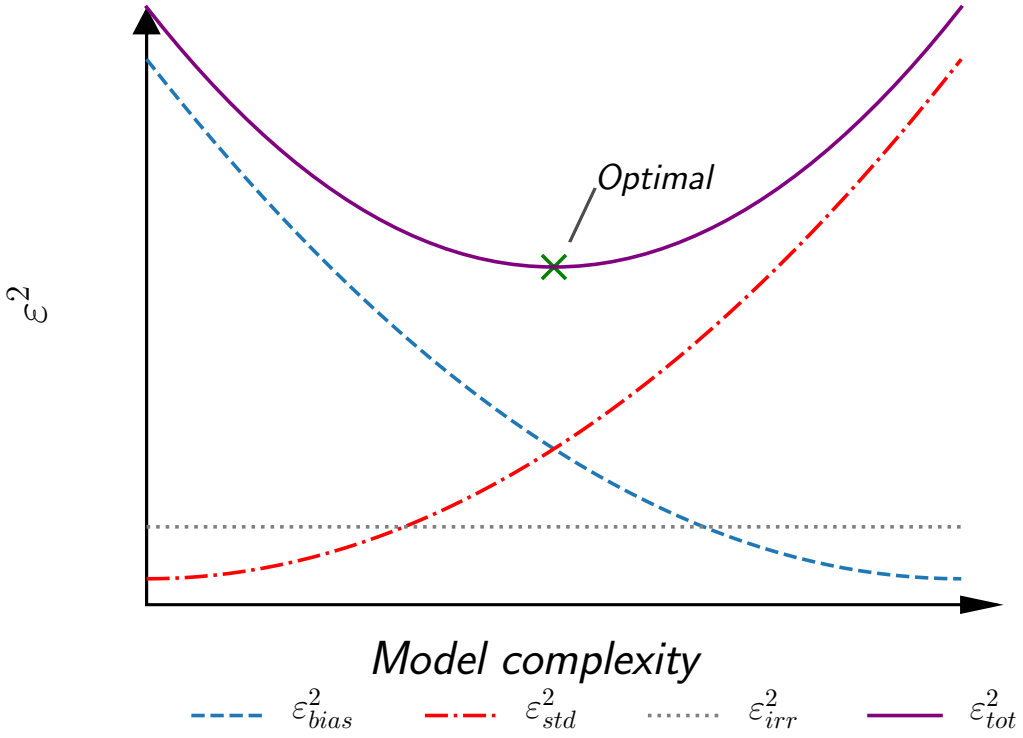


Figure 5.1: The bias variance trade-off.

5.2 Model Definition

The extreme wind speeds relevant for structural design can be expected to lie far in the tail of a probability distribution, which justifies the use of an extreme value distribution to model the relationship between p_t and v_d.

The Gumbel (Extreme Value Type 1) distribution has been widely applied for the prediction of extreme wind speeds, (Kruger *et al.*, 2013*a*; Botha *et al.*, 2018*a*; Palutikof *et al.*, 1999; Hong *et al.*, 2014; JCSS Model Code; Holmes, 2018) it relates the probability of exceedance p to an annual maximum wind speed v as

$$1 - p = \exp\left[-\exp\left\{-\frac{(v - u_v)}{b_v}\right\}\right]. \tag{5.4}$$

The location u_v and scale b_v parameters can be related to the mean μ_v and standard

deviation σ_v of v as

$$\begin{aligned} b_v &= \frac{\sigma_v\sqrt{6}}{\pi} \\ u_v &= \mu_v - \gamma\frac{\sigma_v\sqrt{6}}{\pi}, \end{aligned} \tag{5.5}$$

where $\gamma = 0.577\ldots$ is the Euler Mascheroni constant (Castillo *et al.*, 2005).

Adopting the Gumbel distribution, v_d can be related to p_t using μ_v and σ_v by rearranging Eq. 5.4 and Eq. 5.5 as

$$v_d = \mu_v + \theta_t\sigma_v, \tag{5.6}$$

or in terms of the coefficient of variation $\delta = \sigma/\mu$

$$v_d = \mu_v(1 + \theta_t\delta_v), \tag{5.7}$$

where

$$\begin{aligned} \theta_t &= \theta(p_t) \\ \theta(p) &= -\frac{\sqrt{6}}{\pi}\left[\gamma + \ln(-\ln(1-p))\right]. \end{aligned} \tag{5.8}$$

In this formulation $\theta(p)$ is a standardisation function for a Gumbel distribution and θ_t is the standard score, i.e. the number of standard deviations away from the mean, of the design value. A value of $\theta_t = 5$, corresponding approximately to an annual return period of 1000 years, is considered generally representative of typical loading standard requirements (Table 5.A1).

To estimate the design value, the μ_v and σ_v parameters must themselves be estimated from available wind speed data. Several different modelling approaches relating the σ_v to observed data are available for this purpose and are the primary focus of this study.

5.2.1 Site Statistics Based Design

The μ_v can be estimated from a sample $\{v_1, v_2, \dots v_n\}$ of annual extremes obtained at a specific site as

$$\hat{\mu}_v = \frac{1}{n}\sum_{i=1}^{n} v_i, \tag{5.9}$$

then σ_v can be estimated as

$$\hat{\sigma}_v = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(v_i - \hat{\mu}_v)^2}, \tag{5.10}$$

which can then be used in Eq. 5.6 to estimate v_d

$$\hat{v}_{d:s} = \hat{\mu}_v + \theta_t \hat{\sigma}_v. \tag{5.11}$$

There has been increasing support for this model (*Site Design*), the justification being that if the probability distribution of wind speeds differ geographically, using the recordings at a particular site should give a fairly unbiased estimate of v_d (Ellingwood and Tekie, 1999; Holmes, 2018; Hong *et al.*, 2016).

5.2.2 Regional Statistics Based Design

Generally, samples are recorded at multiple sites within the region for which design wind loads are required. If the region has a fairly homogeneous wind climate, i.e. the probability distributions of wind speeds only differ in magnitude, v_d can be estimated as

$$\hat{v}_{d:r} = \sqrt{\gamma_w}'\hat{\mu}_v. \tag{5.12}$$

In this case $\sqrt{\gamma_w}'$ is a regional frequency factor and is dependant on the probability distribution and target reliability of the region.

For a Gumbel distribution,

$$\sqrt{\gamma_w}' = (1 + \theta_t \overline{\delta}_v) \tag{5.13}$$

where $\overline{\delta}_v$ is the coefficient of variation of wind speeds in the region.

Throughout a perfectly homogeneous region, the coefficient of variation at each site δ_v would be the same, and so site estimates $\hat{\delta}_v = \hat{\sigma_v}/\hat{\mu}_v$ can be aggregated to estimate $\overline{\delta}_v$. This is the key advantage of this model (*Regional Design*), since the variability in an estimate of δ_v is low if observations from many sites are used. Hosking and Wallis (2005) recommends that a set of sample coefficients of variation from N sites $\{\hat{\delta}_{v1}, \hat{\delta}_{v2} \ldots, \hat{\delta}_{vN}\}$, each estimated using n_i observations, be aggregated by weighting with sample size as

$$\hat{\overline{\delta}}_v = \frac{\sum_{i=1}^{N} n_i \hat{\delta}_{vi}}{\sum_{i=1}^{N} n_i} \tag{5.14}$$

to estimate $\overline{\delta}_v$.

A Regional Design is normally referred to as the index flood procedure and has seen widespread adoption in hydrology (Kjeldsen *et al.*, 2002; Parida *et al.*, 1998). Hosking and Wallis (2005) advocate a Regional Design because even if a region is moderately heterogeneous a Regional Design could still yield more accurate fractile estimates than a Site Design, due to the lower model-variance.

There is precedent of a Regional Design being applied to estimate extreme winds (Goel *et al.*, 2004; Hong and Ye, 2014; Modarres, 2008) and was the foundation for various iterations of the Canadian wind loading standard (Hong *et al.*, 2014).

5.2.3 Characteristic Value Based Design

Many wind loading standards follow neither a Site nor Regional Design (Holmes, 2018). Instead, the design is based on characteristic values with a prescribed or intended probability of being exceeded. In the context of wind loading the exceedance probability p_k of the characteristic wind speed v_k is generally referred to using a return period t_k so that

$p_k = 1/t_k$. t_k is often chosen to correspond with the design working life of a structure (SANS 10160-3:2019).

The v_k value is typically based on site estimates and often a Gumbel distribution is assumed (Kruger *et al.*, 2013*b*; Holmes, 2018) so that

$$\begin{aligned} \hat{v}_k &= \hat{\mu}_v + \theta_k \hat{\sigma}_v \\ \theta_k &= \theta(p_k). \end{aligned} \tag{5.15}$$

The v_k value has a high probability of exceedance over the life of the structure (if $t_k = 50$, then $p_{life} = 1 - (1 - 1/50)^{50} = 0.64$). This is unacceptable as a final design value and so v_k is adjusted to a more acceptable probability of exceedance using a partial safety factor $\sqrt{\gamma_w}$.

$$v_{d:k} = \sqrt{\gamma_w} v_k. \tag{5.16}$$

A $\sqrt{\gamma_w}$ is typically a constant for a national standard and is calibrated using regionally representative statistics. If a Gumbel distribution is assumed (Ellingwood and Tekie, 1999; Botha *et al.*, 2018*a*; Hong *et al.*, 2016; Holickỳ, 2009), variability of c neglected, and $\overline{\delta}_v$ used to calibrate $\sqrt{\gamma_w}$, then

$$\sqrt{\gamma_w} = \frac{1 + \theta_t \overline{\delta}_v}{1 + \theta_k \overline{\delta}_v}. \tag{5.17}$$

It should be noted that the partial factor generally given in loading standards also accounts for variability in c. If this is ignored (Cook *et al.*, 2011; Goliger *et al.*, 2017) then, because it is applied after the wind speeds have been squared (Eq. 5.2), the partial factor in a loading standard is equal to γ_w.

This model (*Characteristic Design*) is essentially a combination of a Site and a Regional Design, because the site statistics are used to estimate v_k and $\sqrt{\gamma_w}$ is calibrated using regional statistics. The combination of a Site and Regional Design does not seem to be the original intention of a Characteristic Design, which is related to historic prac-

tice. In most structural design applications a Characteristic Design does not fulfil this combinational role because the location is not correlated with the loads or resistances of a particular structure, for example the concept of a site based design value for the imposed load is illogical since it is dependant on the use of the structure.

The wind load is different because it is dependant on the wind climate at the site of the structure and so a Characteristic Design combines the regional and site statistics. To make this more explicit Eq. 5.15, Eq. 5.16 and Eq. 5.17 can be combined to give

$$\hat{v}_{d:k} = \frac{1 + \theta_t \overline{\delta}_v}{1 + \theta_k \overline{\delta}_v} (\hat{\mu}_v + \theta_k \hat{\sigma}_v). \tag{5.18}$$

Inspection of Eq. 5.18 reveals the balance between a Site and Regional design. If $\theta_k = \theta_t$ then $\sqrt{\gamma_w} = 1$, $\hat{v}_k = \hat{\mu}_v + \theta_t \hat{\sigma}_v$ and Eq. 5.18 simplifies to $\hat{v}_{d:k} = \hat{\mu}_v + \theta_t \hat{\sigma}_v$, which is equivalent to a Site Design. When $\theta_k = 0$ then $\sqrt{\gamma_w} = \sqrt{\gamma_w}'$, $\hat{v}_k = \hat{\mu}_v$ and Eq. 5.18 simplifies to $\hat{v}_{d:k} = (1 + \theta_t \overline{\delta}_v)\hat{\mu}_v$, which is equivalent to a Regional Design. When $0 < \theta_k < \theta_t$ then a Characteristic design results in a design value between a Site and Regional Design.

5.3 Error Quantification

It should be recognised that a probability distribution is a model for the relationship between extreme wind speeds and the frequency of these winds, while a Site, Regional, or Characteristic Design describe the relationship between measured data and the parameters of the distribution. Therefore, the design wind speed v_d can be related to a prediction from either a Site, Regional or Characteristic modelling approach $\hat{v}_{d:a}$, $a \in \{s, r, k\}$ in two steps. First a distribution is selected and some error is introduced in this decision (Chapter 4). Then v_d is estimated by fitting the selected distribution to observations which, because the data available to fit the distribution is limited, introduces more error. The primary focus of this study is on the second step, therefore error originating from the selection of the distribution will be categorised as irreducible.

For the purposes of model selection, irreducible error can be neglected and the relevant MSE associated with a Site, Regional or Characteristic Design can be assessed using a Normalised Root Mean Squared Error (NRMSE) as

$$\varepsilon_a = \frac{\sqrt{\varepsilon^2_{bias:a} + \varepsilon^2_{std:a}}}{v_d}, \tag{5.19}$$

with

$$\varepsilon^2_{bias:a} = \left[E[\hat{v}_{d:a}] - v_d\right]^2, \quad \varepsilon^2_{std:a} = Var[\hat{v}_{d:a}]. \tag{5.20}$$

The normalisation allows $100 \times \varepsilon_a$ to be interpreted as the average percentage error of a prediction associated with a particular modelling approach.

5.3.1 Site Statistics Based Design

Accepting the asymptotic approximations described by Kendall *et al.* (1948), given in Appendix 5.B, the model-bias when using a Site Design is

$$\begin{aligned} \varepsilon^2_{bias:s} &= \left[E[\hat{\mu}_v + \theta_t\hat{\sigma}_v] - (\mu_v + \theta_t\sigma_v)\right]^2 \\ &= \left[\mu_v + \theta_t\sigma_v - \mu_v - \theta_t\sigma_v\right]^2 \\ &= 0. \end{aligned} \tag{5.21}$$

The model-variance when using a Site Design is

$$\begin{aligned} \varepsilon^2_{std:s} =& Var[\hat{\mu}_v + \theta_t\hat{\sigma}_v] \\ =& Var[\hat{\mu}_v] + 2\theta_t Cov[\hat{\mu}_v, \hat{\sigma}_v] + \theta_t^2 Var[\hat{\sigma}_v] \\ =& \frac{\sigma_v^2}{n} + \theta_t\frac{\lambda_v\sigma_v^2}{n} + \theta_t^2\frac{(\kappa_v - 1)}{4n}\sigma_v^2 \\ =& \frac{\mu_v^2\delta_v^2}{n}\left(1 + \lambda_v\theta_t + \frac{(\kappa_v - 1)}{4}\theta_t^2\right). \end{aligned} \tag{5.22}$$

Where λ_v and κ_v are the skewness and kurtosis of the underlying distribution, equal to approximately 1.14 and 5.4 for the Gumbel distribution, and δ_v is the coefficient of

variation at a site.

Substituting Eq. 5.21 and Eq. 5.22 into Eq. 5.19 yields a measure of the error associated with a Site Design

$$\varepsilon_s = \left(\frac{\delta_v}{1+\theta_t\delta_v}\right)\sqrt{\frac{1}{n}\left(1+\lambda_v\theta_t+\frac{(\kappa_v-1)}{4}\theta_t^2\right)}. \tag{5.23}$$

This error is evidently dominated by the model-variance.

5.3.2 Regional Statistics Based Design

The model-bias of a Regional Design is

$$\begin{aligned}\varepsilon^2_{bias:r} &= \left[(1+\theta_t\overline{\delta}_v)E[\hat{\mu}_v]-(\mu_v+\theta_t\sigma_v)\right]^2\\ &=\mu_v^2\theta_t^2(\overline{\delta}_v-\delta_v)^2.\end{aligned} \tag{5.24}$$

The $\varepsilon^2_{bias:r}$ is dependant on how different the δ_v at a site is than the $\overline{\delta}_v$ of the region, i.e. how homogeneous the region is.

Assuming the uncertainty in $\overline{\delta}_v$ is negligible, which would be the case when the number of stations is large ($N \gtrapprox 20$) (Hosking and Wallis, 2005), then the model-variance of a Regional Design is

$$\begin{aligned}\varepsilon^2_{std:r} &= Var[\hat{\mu}_v](1+\theta_t\overline{\delta}_v)^2\\ &=\frac{\sigma_v^2}{n}(1+\theta_t\overline{\delta}_v)^2\\ &=\frac{\mu_v^2\delta_v^2}{n}(1+\theta_t\overline{\delta}_v)^2.\end{aligned} \tag{5.25}$$

Substituting Eq. 5.24 and Eq. 5.25 into Eq. 5.19

$$\begin{aligned}\varepsilon_r &= \frac{\sqrt{\mu_v^2\theta_t^2(\overline{\delta}_v-\delta_v)^2+\frac{\mu_v^2\delta_v^2}{n}(1+\theta_t\overline{\delta}_v)^2}}{\mu_v(1+\theta_t\delta_v)}\\ &=\frac{\sqrt{\theta_t^2(\overline{\delta}_v-\delta_v)^2+\frac{\delta_v^2}{n}(1+\theta_t\overline{\delta}_v)^2}}{1+\theta_t\delta_v}.\end{aligned} \tag{5.26}$$

When comparing the prediction errors of a Site and Regional Design, ε_s is dependant only on the model-variance while a substantial portion of ε_r can be attributed to model-bias.

5.3.3 Characteristic Value Based Design

The model-bias when using Characteristic Design is

$$\begin{aligned}\varepsilon^2_{bias:k} &= \left[\frac{1+\theta_t\overline{\delta}_v}{1+\theta_k\overline{\delta}_v}\left(E[\hat{\mu}_v]+\theta_k E[\hat{\sigma}_v]\right)-\left(\mu_v+\theta_t\sigma_v\right)\right]^2 \\ &= \left[\mu_v\frac{\left(1+\theta_t\overline{\delta}_v\right)\left(1+\theta_k\delta_v\right)-\left(1+\theta_t\delta_v\right)\left(1+\theta_k\overline{\delta}_v\right)}{1+\theta_k\overline{\delta}_v}\right]^2 \\ &= \frac{\mu_v^2(\theta_t-\theta_k)^2(\overline{\delta}_v-\delta_v)^2}{(1+\theta_k\overline{\delta}_v)^2}.\end{aligned} \tag{5.27}$$

The model-variance is

$$\begin{aligned}\varepsilon^2_{std:k} &= \left(\frac{1+\theta_t\overline{\delta}_v}{1+\theta_k\overline{\delta}_v}\right)^2 Var[\hat{\mu}_v+\theta_k\hat{\sigma}_v] \\ &= \left(\frac{1+\theta_t\overline{\delta}_v}{1+\theta_k\overline{\delta}_v}\right)^2 \left(Var[\hat{\mu}_v]+\theta_k^2 Var[\hat{\sigma}_v]+2\theta_k Cov[\hat{\mu}_v,\hat{\sigma}_v]\right) \\ &= \frac{\mu_v^2\delta_v^2(1+\theta_t\overline{\delta}_v)^2\left(1+\lambda_v\theta_k+\frac{(\kappa_v-1)}{4}\theta_k^2\right)}{(1+\theta_k\overline{\delta}_v)^2 n}.\end{aligned} \tag{5.28}$$

Substituting Eq. 5.27 and Eq. 5.28 into Eq. 5.19

$$\varepsilon_k = \frac{\sqrt{(\theta_t-\theta_k)^2(\overline{\delta}_v-\delta_v)^2+\frac{\delta_v^2}{n}(1+\theta_t\overline{\delta}_v)^2\left(1+\lambda_v\theta_k+\frac{(\kappa_v-1)}{4}\theta_k^2\right)}}{(1+\theta_k\overline{\delta}_v)(1+\theta_t\overline{\delta}_v)}. \tag{5.29}$$

Inspection of the expressions for the error of a Characteristic Design (Eq. 5.27, Eq. 5.28, and Eq. 5.29) reveal that if $\theta_k = \theta_t$ then they simplify to the expressions for the error of a Site Design (Eq. 5.21, Eq. 5.22, and Eq. 5.23). Similarly if $\theta_k = 0$ then they simplify to the expressions for the error of a Regional Design (Eq. 5.24, Eq. 5.25, and Eq. 5.26).

5.4 Sources of Error

5.4.1 Model Variance

The ε^2_{std} expressions (Eq. 5.22, Eq. 5.25, Eq. 5.28), show that the model-variance is proportional to δ_v. This is expected as a more variable wind climate would increase the uncertainty of predictions.

The ε^2_{std} is proportional to $1/n$ therefore, MSE decreases as n increases, but the rate of the decrease slows as n increases. This suggests that focus on extending records of stations for which data is limited could be prioritised over established stations with long records.

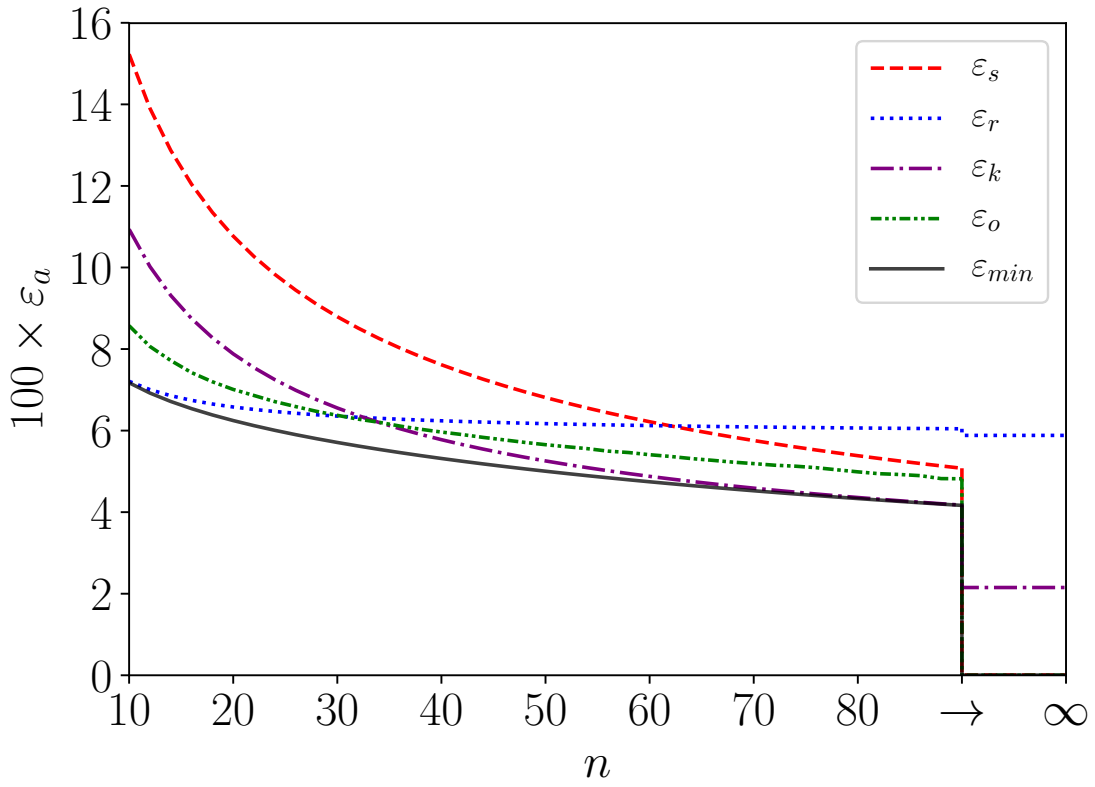


Figure 5.2: % ε_a vs n for $\theta_t = 5$, $\delta_v = 0.14$ and $\overline{\delta}_v = 0.12$. Note ε_o and ε_{min} are defined in Section 5.5.

The ε_a for typical n values (Holmes, 2018) are shown in Figure 5.2. Comparing ε_s with ε_r indicates that a Regional Design is preferred when n is low. As n increases the

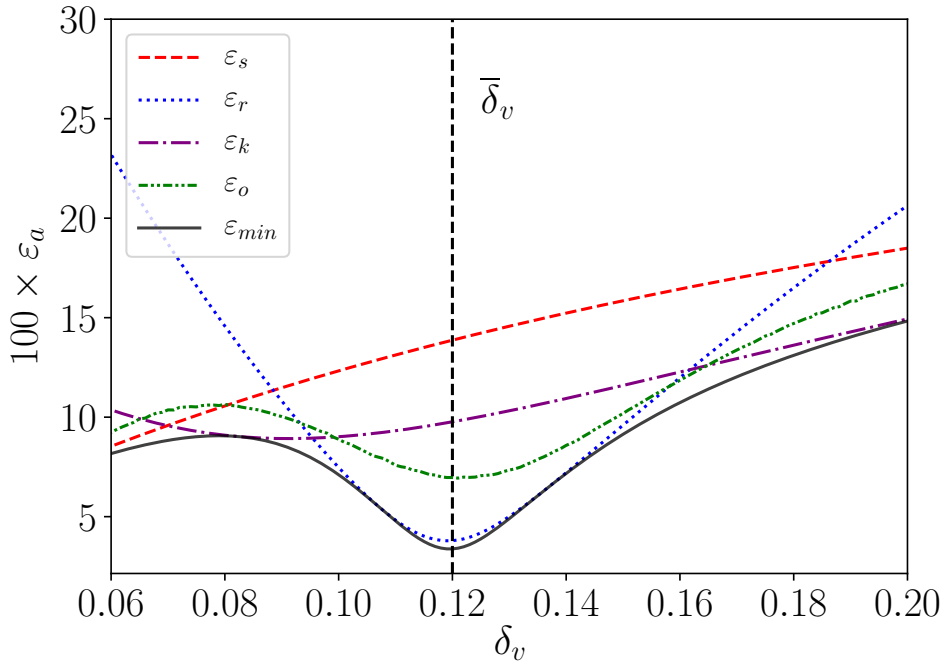
ε^2_{std} decreases and so the performance of a Site Design improves until it is preferred to a Regional design. This is because $\varepsilon_s \sim \varepsilon^2_{std:s}$ and so as $n \to \infty$, then $\varepsilon_s \to 0$. This is not the case for a Regional or Characteristic Design for which the decrease in error is limited by model-bias.

5.4.2 Model Bias

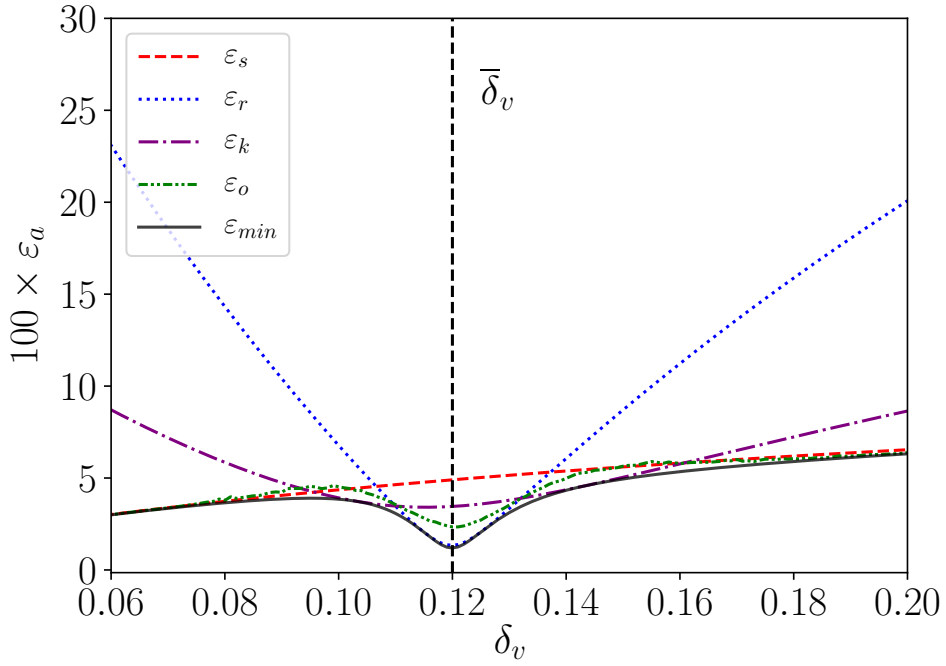
Unlike a Site Design, a significant portion of the error of a Regional Design can be attributed to model-bias (Eq. 5.24). This model-bias is low when δ_v is near $\overline{\delta}_v$ i.e. $(\overline{\delta}_v - \delta_v)^2$ is small because the model-bias, which dominates a Regional Design, is dependant on the difference between δ_v and $\overline{\delta}_v$. The δ_v represents variability of the wind speed at a specific site and $\overline{\delta}_v$ represents the average site variation of the wind speed over a region. Thus $(\overline{\delta}_v - \delta_v)^2$ is an indicator of the homogeneity of the wind climate.

The errors obtained for the various models ε_a are shown in Figure 5.3, where the range of δ_v and $\overline{\delta}_v = 0.12$ can be considered typical (JCSS Model Code; Hong *et al.*, 2016; Holmes, 2018). This confirms that a Regional Design is preferred when model-bias $(\overline{\delta}_v - \delta_v)^2$ is low. Comparing Figures 5.3a and 5.3b, the range of δ_v values for which a Regional Design is preferred over a Site Design is greater for a lower n.

Hosking and Wallis (2005) stress the importance of homogeneity for a good Regional Design, and so a substantial effort is made to separate regions that cannot be considered homogeneous. Kruger *et al.* (2011) found that South Africa was dominated by synoptic scale winds along the coast and south-west of the country, while the interior was dominated by convective thunderstorms. Botha (2016) found a significant difference in the mean δ_v values of these two climatic areas, which should therefore be considered separate regions with distinct $\overline{\delta}_v$ values. Provided a $\hat{v}_d$ can be expressed in terms of a set of parameters for each mechanism, the procedure described in Section 2.3.2 could be applied to account for independent mechanisms.

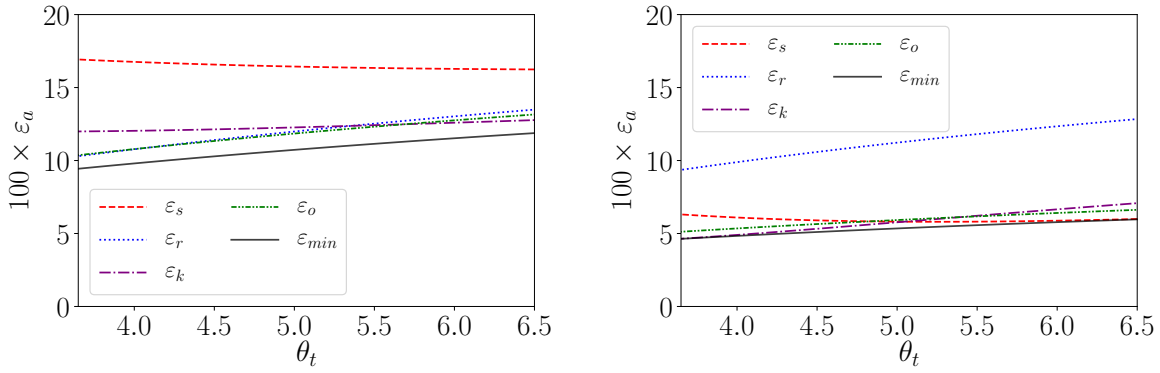


(a) $n = 10$



(b) $n = 80$

Figure 5.3: ε_a vs. δ_v when $\overline{\delta}_v = 0.12$ and $\theta_t = 5$. Note ε_o and ε_{min} are defined in Section 5.5.



(a) $n = 10$ (b) $n = 80$

Figure 5.4: ε_a vs. θ_t, for $\delta_v = 0.16$ and $\overline{\delta}_v = 0.12$. Note ε_o and ε_{min} are defined in Section 5.5.

5.4.3 Model Input

The higher θ_t, the greater the magnitude of the error for any design choice because, regardless of the model, predictions further into the tail of a distribution are more uncertain. The normalisation in ε_a separates the error from the magnitude and allows for the relative performance across a range of θ_t values to be compared. This relative performance is shown (Figure 5.4) to be fairly consistent across a large range θ_t.

The relative performance of a Site, Regional or Characteristic Design is more affected by other parameters, because both the model-bias and the model-variance increase with θ_t. This is demonstrated in Figure 5.4 where $\hat{v}_{d:r}$ is preferred when n is low (Figure 5.4a) and $\hat{v}_{d:s}$ is preferred when n is high (Figure 5.4b).

5.4.4 Irreducible Error

Within the framework of this study, as described in Section 5.2, sources of irreducible error include measurement error and the choice of distribution, neither of which is truly irreducible.

Where long records of wind data are available, measurement errors in older records have been shown to be substantial (Kruger *et al.*, 2011; Miller *et al.*, 2013). When possible these records should be corrected, as attempted in Chapter 3.

The error in the distribution choice could be reduced by choosing a more appropriate distribution (Holmes, 2018; Palutikof *et al.*, 1999; Cook and Harris, 2004; Walshaw, 1994; Rózsás and Sỳkora, 2016). The Gumbel distribution is considered in this study due to its widespread use in codification efforts, (Kruger *et al.*, 2013*a*; Botha *et al.*, 2018*a*; Hong *et al.*, 2014; JCSS Model Code; Gulvanessian *et al.*, 2012) and because it's simplicity allows the development of the expressions in Section 5.2 and 5.3, which yields a clearer understanding of the problems associated with representing the wind climate in codification.

The Gumbel distribution represents one asymptotic form of the Generalised Extreme Value distribution. The primary drawback of using asymptotic distributions is that a large portion of recorded extremes are discarded since only a single value per block is used. Other approaches such as the Peak Over Thresholds or Method of Independent Storms do not suffer from this disadvantage. These approaches utilise more data at a given site and so the model-variance decreases, and the relative performance of a site design should improve significantly. The drawback of these approaches is that separation from other wind mechanisms and decisions such as the threshold selection introduce new sources of error (Palutikof *et al.*, 1999). It also means that more labour is required to process data at each site.

Asymptotic theory can be justified for practical reasons in some cases. Both Chapter 4 and Hong *et al.* (2014) argue that if quality control of data is a substantial constraint, then restricting the analysis to annual extremes limits the effort of data extraction and allows data from more stations to be considered. This should more accurately capture the extreme wind climate over the large region required for codification.

The Gumbel distribution is shown to be the preferred asymptotic form, despite the inflexibility and potential for model-bias associated with fixing the shape parameter, this was shown in Chapter 4 for the extended dataset from Chapter 3 and has also been shown for a number of other datasets (Hong *et al.*, 2014; Baravalle and Köhler, 2018). Chapter 4 demonstrates that the overall model-bias of the Gumbel distribution can be substantially decreased by regionally preconditioning the wind speed, in which case the expressions

above can refer to the statistics of the preconditioned wind speed.

5.4.5 Model Complexity

A Site Design involves the estimation of variation σ_v for each site within a region and so is a comparatively complex model with many parameters. A Site Design tends to perform well when n and $(\overline{\delta}_v - \delta_v)^2$ are high, i.e. when the potential for model-variance is low and model-bias high, because a Site Design is a low bias high variance model.

A Regional Design is substantially less complicated with only a single estimate of the variation $\overline{\delta}_v$ required for the entire region. A Regional Design tends to performs well when n and $(\overline{\delta}_v - \delta_v)^2$ are low, i.e. when the potential for model-variance is high and model-bias low, because a Regional design is a high bias low variance model.

A Characteristic Design is transitional between a Site and Regional Design with the weighting controlled by θ_k. Therefore, θ_k can be considered a parameter controlling the model complexity. The most common nominal return period found in standards is 50-years ($\theta_k = 2.59$) resulting in a relatively even balance between a Site and Regional Design for a typical $\theta_t \approx 5$. The ε_k for this case is shown in Figures 5.2, 5.3 and 5.4. It appears to generally perform well by moderating the high error when a Site or Regional Design performs poorly.

The choice of θ_k seems to be fairly arbitrary with various return periods employed by different national standards (Holmes, 2018). To investigate the relationship between θ_k and MSE, the θ_k was varied for a typical case ($\mu_v = 25$ m/s, $n = 20$, $\delta_v = 0.18$, $\overline{\delta}_v = 0.12$) and the MSE calculated using Eq. 5.27 and Eq. 5.28. The results are displayed in Figure 5.5.

Inspection of Figure 5.5 shows that $\varepsilon^2_{bias:k}$ has a minimum value corresponding to some θ_k. This minimum can be found by solving θ_k for

$$\frac{\partial \varepsilon^2_{bias:k}}{\partial \theta_k} = 0, \tag{5.30}$$

which yields

$$\underset{\theta_k}{\text{argmin}}[\varepsilon^2_{bias:k}] = \theta_t. \tag{5.31}$$

Which is unsurprising since $\theta_k = \theta_t$ is equivalent to a Site Design, for which $\varepsilon^2_{bias:k} = 0$. This confirms that a Site Design is a low bias design.

Figure 5.5 also shows that $\varepsilon^2_{std:k}$ has a minimum value corresponding to some θ_k. Following the same procedure as before with $\varepsilon^2_{std:k}$ gives

$$\underset{\theta_k}{\text{argmin}}\,[\varepsilon^2_{std:k}] = \frac{2\alpha - 4\overline{\delta}_v}{2\overline{\delta}_v\alpha - \kappa + 1}. \tag{5.32}$$

For typical values of $\overline{\delta}_v$ this $\underset{\theta_k}{\text{argmin}}[\varepsilon^2_{std:k}]$ is near 0 ($\overline{\delta}_v = 0.1 \Rightarrow \underset{\theta_k}{\text{argmin}}[\varepsilon^2_{std:k}] = -0.45$, $\overline{\delta}_v = 0.2 \Rightarrow \underset{\theta_k}{\text{argmin}}[\varepsilon^2_{std:k}] = -0.38$), which confirms that a Regional Design ($\theta_k = 0$) is a low variance design. It should be recognised that $\theta_k = 0$ for a Regional Design is somewhat arbitrary and originates from the choice to use μ_v as a reference value.

In Figure 5.5 when $\underset{\theta_k}{\text{argmin}}[\varepsilon^2_{std:k}] < \theta_k < \underset{\theta_k}{\text{argmin}}[\varepsilon^2_{bias:k}]$ then $\varepsilon^2_{std:k}$ is an increasing function of θ_k and $\varepsilon^2_{bias:k}$ is a decreasing function of θ_k. It follows that between $\underset{\theta_k}{\text{argmin}}[\varepsilon^2_{std:k}]$ and $\underset{\theta_k}{\text{argmin}}[\varepsilon^2_{bias:k}]$ there is a trade-off between the model-bias and model-variance with some optimal combination, corresponding to an optimal return period $\theta_{k:opt}$, which minimises the MSE. Outside of these limits, the model-bias and model-variance are both either increasing or decreasing, and so will always result in greater error.

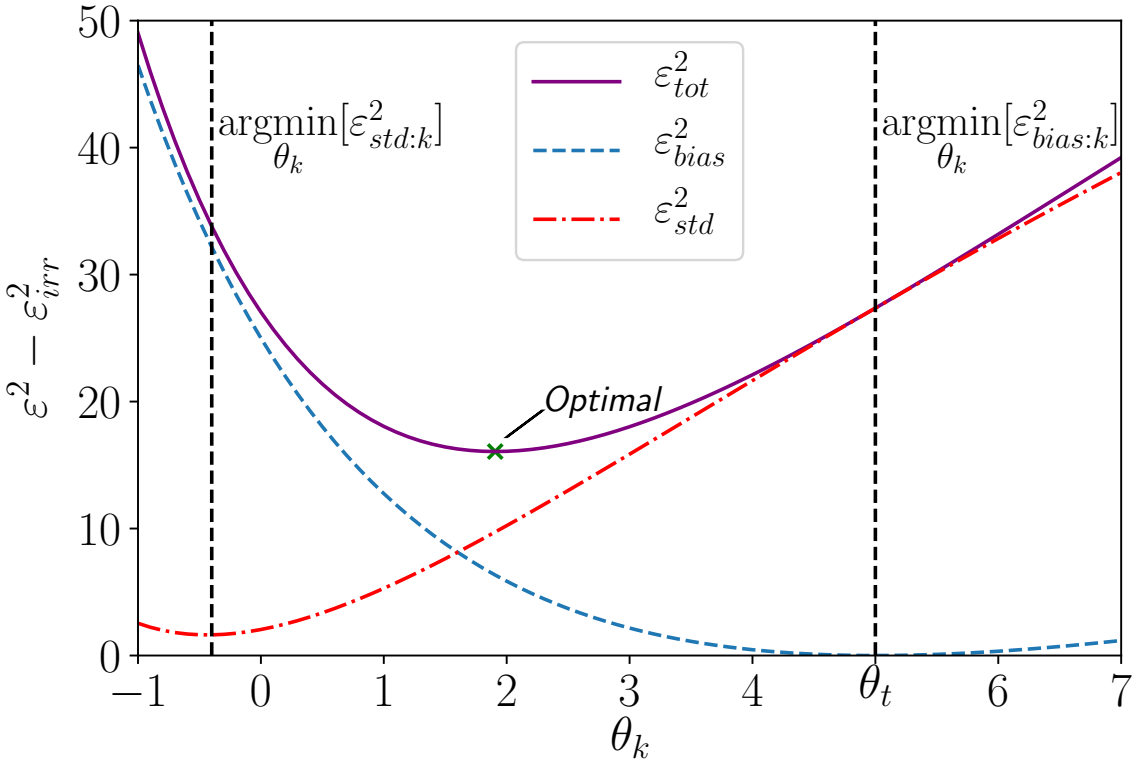


Figure 5.5: ε^2 vs. θ_k for $\mu_v = 25$ m/s, $\delta_v = 0.16$, $n = 20$ and $\overline{\delta}_v = 0.12$.

5.5 Minimum Mean Squared Error

A Characteristic Design ($\theta_k \in \mathbb{R}$) is more flexible than a Site Design ($\theta_k = \theta_t$) or Regional Design ($\theta_k = 0$). Therefore, it should generally be able to yield a more optimal/lower MSE design than either a Site or Regional Design.

5.5.1 Development

Inspection of Eq. 5.28 shows that model-variance decreases as n increases, while Eq. 5.27 indicates that the model-bias remains unchanged. Since the portion of the MSE attributed to the model-variance has decreased, model-bias plays a bigger role and the optimal model shifts nearer to a Site Design. This concept is demonstrated in Figure 5.6a which suggests $\theta_{k:opt} \to \theta_t$ as n increases.

Inspection of Eq. 5.27 shows that model-bias decreases as $(\overline{\delta}_v - \delta_v)^2$ decreases, and so because a greater portion of the MSE is attributed to model-variance the optimal model

shifts nearer to a Regional Design. This concept is demonstrated in Figure 5.6b where $\theta_{k:opt}$ decreases as $(\overline{\delta}_v - \delta_v)^2$ decreases.

Figure 5.6 implies that $\theta_{k:opt}$ is dependant on the quantity of data at a site and the conditions of the site relative to the region. This is more clearly defined with an expression for $\theta_{k:opt}$, which can be found by minimising the MSE by solving for θ_k in

$$\frac{\partial \varepsilon_k^2}{\partial \theta_k} = 0, \tag{5.33}$$

which yields

$$\theta_{k:opt} = \frac{\delta_v^2\left(1+\theta_t\overline{\delta}_v\right)\left(2\alpha-4\overline{\delta}_v\right) - 4(\overline{\delta}_v-\delta_v)^2 n\theta_t}{\delta_v^2\left(1+\theta_t\overline{\delta}_v\right)\left(2\overline{\delta}_v\alpha-\kappa+1\right) - 4(\overline{\delta}_v-\delta_v)^2 n}. \tag{5.34}$$

This confirms the upper limit observed in Figure 5.6 with

$$\lim_{n\to\infty}\theta_{k:opt} = \lim_{(\overline{\delta}_v-\delta_v)^2\to\infty}\theta_{k:opt} = \theta_t. \tag{5.35}$$

Proving that when the quantity of data at a site or the heterogeneity of the region is high, the optimal design approaches a Site Design.

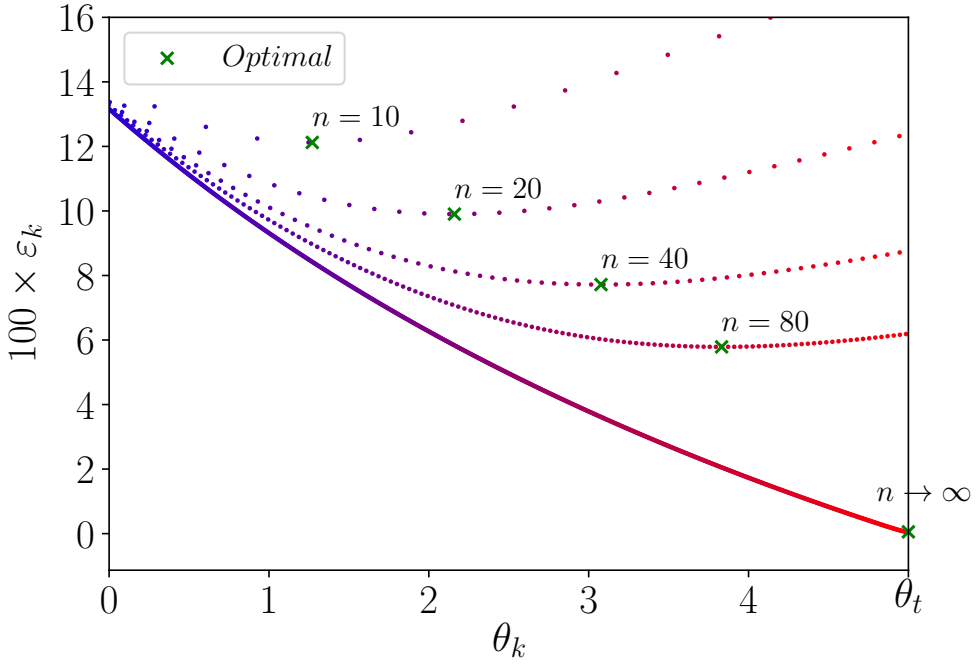
It also confirms the lower limit observed in Figure 5.6 with

$$\lim_{n\to 0}\theta_{k:opt} = \lim_{(\overline{\delta}_v-\delta_v)^2\to 0}\theta_{k:opt} = \frac{2\alpha-4\overline{\delta}_v}{2\overline{\delta}_v\alpha-\kappa+1}. \tag{5.36}$$

Proving that when the quantity of data at a site or the heterogeneity of the region is low, $\theta_{k:opt}$ approaches a minimum model-variance design which tends to be near a Regional Design.

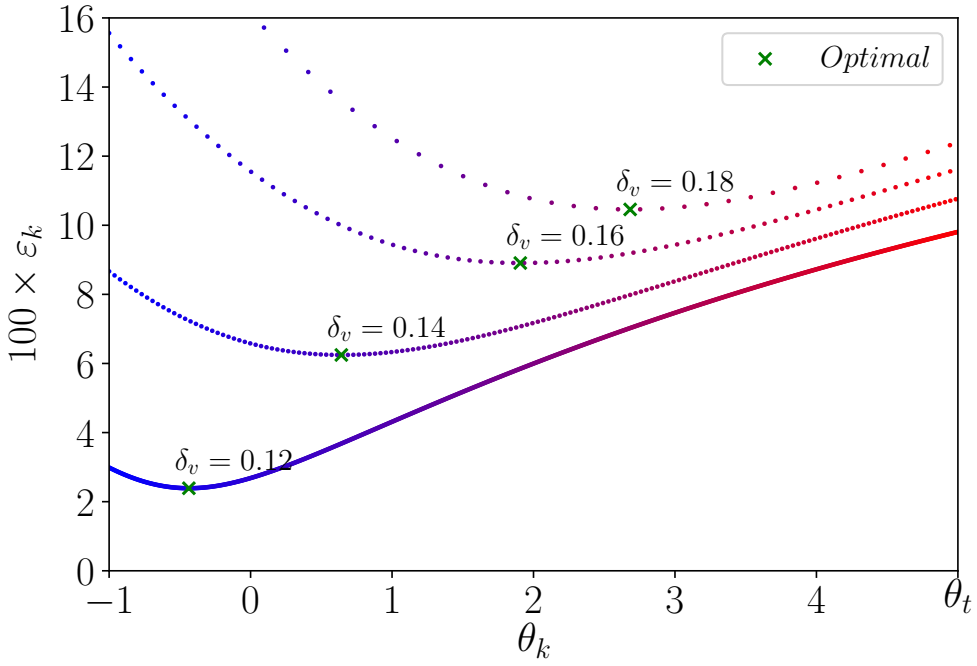
While the representation of $\theta_{k:opt}$ with Eq. 5.34 helps define when a Site or Regional design could be considered, it also presents an opportunity to develop an alternative minimised MSE model (*Optimised Design*) by substituting $\theta_{k:opt}$ in Eq. 5.18

$$\hat{v}_{d:o} = \frac{1+\theta_t\overline{\delta}_v}{1+\theta_{k:opt}\overline{\delta}_v}(\hat{\mu}_v + \theta_{k:opt}\hat{\sigma}_v). \tag{5.37}$$



(a) $n = \{10, 20, 40, 80, \infty\}$, $\delta_v = 0.18$, $\overline{\delta}_v = 0.12$

.



(b) $\delta_v = \{0.18, 0.16, 0.14, 0.12\}$, $n = 20$, $\overline{\delta}_v = 0.12$

.

Figure 5.6: ε_k vs. θ_k, with $\theta_t = 5$

5.5.2 Error Quantification

An Optimised Design should yield the minimum possible MSE within the constraints defined in Section 5.2. This error ε_{min} can be found by first using Eq. 5.34 to calculate $\theta_{k:opt}$, then substituting this into Eq. 5.29 to find $\varepsilon_{min} = \varepsilon_k(\theta_{k:opt})$, and is shown in Figures 5.2, 5.3 and 5.4 for a variety of cases. Which demonstrates how an Optimised design shifts between a Site and Regional Design to minimise error.

The issue with achieving ε_{min} is that δ_v is unknown and required to calculate $\theta_{k:opt}$. This can be overcome by using the sample coefficient of variation $\hat{\delta}_v$ to estimate the optimal return period $\hat{\theta}_{k:opt}$, which means the MSE that is practically realised for an Optimised Design ε_o is greater than ε_{min} due to additional parameter uncertainty.

The ε_o could be evaluated by finding the expected value of the error when $\theta_{k:opt}$ is estimated using $\hat{\delta}_v$

$$\varepsilon_o = \int_{-\infty}^{\infty} \varepsilon_k(\hat{\theta}_{k:opt}) f(\hat{\delta}_v|\delta_v, n) d\hat{\delta}_v, \tag{5.38}$$

where $f(\hat{\delta}_v|\delta_v, n)$ is the probability density function of $\hat{\delta}_v$ for a given δ_v and n. Equation 5.38 can be computed using simulation for δ_v, n, θ_t and $\overline{\delta}_v$. First, m samples each of length n are drawn from a Gumbel distribution with coefficient of variation δ_v. Then the sample coefficient of variation $\hat{\delta}_v$ for each simulated sample is calculated $\{\hat{\delta}_{v1}, \hat{\delta}_{v2}, \ldots, \hat{\delta}_{vm}\}$ and, provided m is large enough, used to find ε_o

$$\varepsilon_o = \frac{1}{m} \sum_{i=1}^{m} \varepsilon_k(\hat{\theta}_{k:opt}|\hat{\delta}_{vi}, \delta_v, n, \theta_t, \overline{\delta}_v). \tag{5.39}$$

This procedure was used to find ε_o for a variety of cases shown in Figures 5.2, 5.3 and 5.4, where an $m = 10\ 000$ was used. The effect of having to estimate δ_v can be observed in the difference between ε_o and ε_{min}. Despite this additional parameter uncertainty, an Optimised Design appears to still offer consistent performance across a large range of n and δ_v values.

5.6 Application of Modelling Approaches to Stations in South Africa

Botha *et al.* (2018*b*) analysed 76 stations throughout South Africa and used a hierarchical network to support a coefficient of variation equal to 0.31 as representative of extreme wind pressure for the country (for wind speeds $\overline{\delta}_v \approx 0.15$). This was then used to calibrate the partial factor of 1.6 for wind loading in the South African standard (Botha *et al.*, 2018*b*) ($\gamma_w = 1.6$), which uses a 50-year return period characteristic value ($\theta_k = 2.59$).

Applying this information in the framework developed in Section 5.2.3, the θ_t currently provided by the Characteristic Design approach in the South African standard (SANS 10160-3:2019) can then be back-calculated with Eq. 5.18 and found to be $\theta_t \approx 5$ for the reference level of reliability (Retief and Dunaiski, 2009). This is higher than the $\theta_t = 4.3$ in Table 5.A1, likely because the uncertainty in c_k was neglected in the derivation of Table 5.A1.

Given $\theta_t = 5$, $c_k = 1$, $\rho = 1.25$ kg/m^3 (Holickỳ, 2009) and wind speed data, the predicted design wind pressure for a Site $\hat{q}_{d:s}$, Regional $\hat{q}_{d:r}$, Characteristic $\hat{q}_{d:k}$ and Optimised design $\hat{q}_{d:o}$ can be calculated (Eq. 5.2). This allows the practical implications of a particular modelling approach to be compared. Table 5.1 shows the site statistics and design wind pressure predictions for several stations in South Africa.

Table 5.1: Models applied to South African stations: $\theta_t = 5$, $\overline{\delta}_v = 0.15$. Design pressures are given in Pascals.

Station	n	$\hat{\mu}_v$	$\hat{\delta}_v$ $[CI_{90\%}]$	$q_{d:s}$	$q_{d:r}$	$q_{d:k}$	$\hat{\theta}_{k:opt}$	$q_{d:o}$
Wonderboom	10	29.7	0.14 [0.09,0.24]	1584	1693	1621	-0.21	1701
Johannesburg	61	27.3	0.13 [0.12,0.18]	1250	1424	1309	2.32	1318
Bird Island	11	26.7	0.07 [0.05,0.12]	788	1362	969	4.50	818

Inspection of Table 5.1 shows how $\hat{q}_{d:s}$ varies more than $\hat{q}_{d:r}$, because they each use their own $\hat{\delta}_v$ rather than all using $\overline{\delta}_v$. The $\hat{q}_{d:k}$ predictions are around midway between $\hat{q}_{d:s}$ and $\hat{q}_{d:r}$ because $\theta_k = 2.59$ is around midway between 0 and θ_t. The $\hat{q}_{d:o}$ predictions are dependent on more information than $\hat{q}_{d:s}$, $\hat{q}_{d:r}$ or $\hat{q}_{d:k}$.

The $(\overline{\delta}_v - \hat{\delta}_v)^2$ and n are fairly low for the Wonderboom station, indicating that there is little evidence that the station is significantly different to the regional average. This is reflected in the low $\hat{\theta}_{k:opt}$ which applies more weight to a Regional Design and results in a $\hat{q}_{d:o}$ near $\hat{q}_{d:r}$. In fact, the $\hat{\theta}_{k:opt} < 0$ and so even more weight is applied to the regional average coefficient of variation than in a Regional Design. At the Johannesburg station there is substantially more data $n = 61$ and so, despite a similar δ_v as the Wonderboom station, a higher $\hat{\theta}_{k:opt}$ is obtained and more weight is applied to a Site Design. This results in a $\hat{q}_{d:o}$ near $\hat{q}_{d:k}$.

A low n does not necessarily mean that a Site Design is inappropriate. This is demonstrated by the Bird Island station which is located on a small island off the Southern Coast of the country, far from the other two stations. The wind climate on this island is likely to be different to the rest of the country leading to an anomalously low $\hat{\delta}_v$ and high $(\overline{\delta}_v - \hat{\delta}_v)^2$. This shifts the balance towards a Site Design, with a high $\hat{\theta}_{k:opt}$ resulting in a $\hat{q}_{d:o}$ near $\hat{q}_{d:s}$.

5.7 Conclusion

To ensure that a structure designed according to a standard performs adequately during an extreme wind event, a model relating the design wind speed to a target probability of exceedance is required.

If the Gumbel distribution is accepted as a reasonable model, then an estimate of the variability using measured wind speed data is needed. Spatial variation of the wind climate means that it is unclear which data are relevant at a particular location to make such an estimate. Several modelling approaches for incorporating measured data were compared by developing expressions and quantifying the errors associated with each modelling approach.

Besides providing a basis for model selection, explicitly representing the error provides clarity on the issues involved with different modelling approaches. It is shown that the error attributed to model-variance is proportional to $1/n$ which is used to argue that

stations where data is limited should be prioritised over established stations with long records. It is also shown why a single partial factor can be questioned when regional differences in the wind climate exist, as is the case in South Africa (Botha, 2016; Kruger *et al.*, 2010).

Initially three modelling approaches were considered: Site, Regional, and Characteristic-value based Design. A Site Design involves using only the data measured at a particular location to estimate a design fractile. This approach has the advantage of being unbiased, but requires fairly long records to avoid high model-variance predictions. A Regional Design aggregates the variability estimates from many sites in a region. It has the advantage of decreasing the variability of predictions, but estimates derived using this method could be substantially biased if the wind climate is not homogeneous, i.e the regional aggregates are not representative of the variability for a particular site.

A Characteristic Design is commonly used in design standards and involves estimating a lower fractile value as a characteristic value using site data and then using a partial factor, that has been calibrated using regional aggregates, to obtain the design value. In this case a Characteristic Design can be understood as a combination of a Site Design (low bias) and a Regional Design (low model-variance). By moderating the errors associated with unfavourable situations for these two design options a Characteristic Design with a moderate return period is shown to offer more consistent performance than a Site or Regional Design.

Because there is a trade-off between the model-bias and model-variance, the best model is typically not a high variance or high bias model, but rather is related to an optimal level of complexity. By varying the return period of the characteristic value, the weight place on the comparatively complex Site or simpler Regional Design can be controlled. This insight is used to develop an expression for a minimum mean squared error estimate of the design value (Optimised Design).

An Optimised Design is shown, despite additional parameter uncertainty, to give consistent performance by adapting the weighting between a Site and Regional Design depending on the information at a particular site. This is demonstrated using data from

several stations in South Africa, where an Optimised Design varies based on the amount of data at the station and how that data corresponds with the regional aggregate.

Appendix 5

5.A Reliability Requirements for Design Wind Speed

Target probability p_t is referred to using the target reliability index $\beta_t = \Phi^{-1}(1 - p_t)$, with Φ the standard cumulative normal distribution function, and differs depending on the intended purpose of the structure and the associated reliability class (Retief and Dunaiski, 2009). Separation of the target reliability of a load can be done by either assuming a FORM sensitivity factor of -0.7 (EN 1990; Goliger *et al.*, 2017) or by back calculating p_t from existing provisions (Cook *et al.*, 2011).

ASCE 7-16 and ISO 4354:2009 directly provide the wind speed annual p_t for a reliability class as a return period $t_t = 1/p_t$. For standards that do not explicitly give p_t it can be computed as $p_t = \Phi(-0.7\beta_t)$. The return period t_t and standard score θ_t for wind speeds from a number of standards are shown in Table 5.A1. ISO, ASCE and SANS generally have similar reliability targets.

Table 5.A1: Annual reliability requirement for wind action from different standards in terms of return period t_t and standard score θ_t. Note that the ISO importance levels and ASCE consequence classes are presented as equivalent to the SANS reliability classes.

Reliability	ISO:4354		ASCE:7-16		SANS:10160-3	
Class	t_t	θ_t	t_t	θ_t	t_t	θ_t
1	200	3.7	300	4	193	3.7
2	500	4.4	700	4.7	425	4.3
3	1000	4.9	1700	5.3	1044	5.0
4	2000	5.5	3000	5.8	2876	5.8

5.B Expected Values and Variances of Site Statistics

The sample mean $\hat{\mu}_v$ is normally distributed (central limit theorem) with

$$E[\hat{\mu}_v] = \mu_v \text{ and } Var[\hat{\mu}_v] = \frac{\sigma_v^2}{n}. \tag{5.B1}$$

The expected value and variance of the sample standard deviation $\hat{\sigma}_v$ is approximately (Kendall *et al.*, 1948)

$$E[\hat{\sigma}_v] = \sigma_v \text{ and } Var[\hat{\sigma}_v] = \frac{(\kappa_v - 1)}{4n}\sigma_v^2, \tag{5.B2}$$

where κ_v is the kurtosis of the underlying distribution. For skew distributions $\hat{\mu}_v$ and $\hat{\sigma}_v$ are correlated. The covariance can be shown (Kendall *et al.*, 1948) to be approximately

$$Cov[\hat{\mu}_v, \hat{\sigma}_v] = \frac{\lambda_v \sigma_v^2}{2n}. \tag{5.B3}$$

where λ_v is the skewness of the distribution.

The above expressions are only asymptotic approximations and are more accurate for increasing n. Monte Carlo simulation was used to confirm that the modelling approach errors based on these approximations (Section 5.3) are still sufficiently accurate.

Chapter 6

Updating the South African Wind Load Formulation

FP Bakker, N de Koker, C Viljoen

Intended for submission to the *Journal of the South African Institution of Civil Engineering*.

Summary

The South African wind loading standard (SANS 10160-3:2019) recently adopted an improved map of characteristic basic wind speeds and increased the wind loading partial safety factor from 1.3 to 1.6. These changes represent an overhaul of the design wind loads throughout South Africa and were the result of a number of studies on the wind loading standard. Since these studies were conducted, substantially more wind speed data has been made available. This investigation aims to use this data to assess the current design loads by estimating location specific design values, that maintain the current reliability level of the standard. A statistical test is developed to asses whether the design values in SANS 10160-3:2019 can be supported by the new data. It is found that a number of updates could be considered. These are incorporated into a new recommended map of

basic wind speeds that could be considered for inclusion in the next revision of SANS 10160-3.

Notation

v_{50}	Characteristic wind speeds in SANS 10160-3 (50-year return period)
ρ	Air density
c_k	Wind load modification factors in SANS 10160-3
γ_w	Wind load partial safety factor
l	Performance function
r	Random variable representing resistance
g	Random variable representing permanent loading
q	Random variable representing wind pressure
c	Random variable representing wind load modification factors
v	Random variable representing wind speed
β_t	Target reliability index
β_q	Exceedance probability of the design wind load in terms of reliability index
α	Sensitivity factor
Φ	Standard normal CDF
v_d	Design wind speed
p	Exceedance probability
p_t	Target exceedance probability
$p_{t:v_d}$	Exceedance probability of design wind speed
c_{prob}	Return period modification factor
c_d	Design value of wind load modification factor
$v^{\overline{w}}$	Regionally preconditioned wind speed
$\mu_{v^{\overline{w}}}$	Mean of the regionally preconditioned wind speed
$\sigma_{v^{\overline{w}}}$	Standard Deviation of the regionally preconditioned wind speed
θ	Standardisation function
n	Number of available annual maxima in a series
λ	Skewness
κ	Kurtosis
N	Number of available stations
f	Generic function
Θ	Set of parameters
$\hat{q}_d$	Estimated design wind pressure
$q_{d:sans}$	Design wind pressure in SANS 10160-3
Δ_{sans}	Normalised difference between estimated and SANS design values.
H_0	Null hypothesis
v_{sim}	Simulated wind speeds
$v_{50:updated}$	Updated 50-year wind speed values
v_k	Characteristic wind speed of unspecified return period
δ_v	Wind speed coefficient of variation
z	Transformed random variable
x	Generic random variable

6.1 Introduction

The current South African wind loading standard (SANS 10160-3:2019) is largely based on the Eurocode (EN 1991-1-4), and uses Davenport (1982)'s wind loading chain to convert a representative wind speed v_{50} to a design wind load/dynamic pressure as

$$\begin{aligned} q_{d:sans} &= \frac{1}{2}\rho\gamma_w c_k v_{50}^2 \\ c_k &= \prod_{i=1}^{m} c_{ik}, \end{aligned} \tag{6.1}$$

where ρ is the air density. The v_{50} value is the characteristic value of the basic wind speed, with an annual exceedance probability $p = 0.02$, representing wind speed measured at 10 metres above ground level in open country terrain with low vegetation. The load is tailored to site and structural conditions using c_k, which is the product of other factors c_{ki}. These account for different aspects of the particular load, including the change in load due to height, topography and aerodynamics of the structure. To ensure that the calculated load meets the reliability requirements of the standard (SANS 10160-1:2019), uncertainties are accounted for using the partial safety factor γ_w.

New wind speed data has recently been made available (Chapter 3) and more research into applicable statistical techniques has been conducted (Chapter 4, Chapter 5), which present an opportunity to improve the representation of the South African wind climate in SANS 10160-3.

6.1.1 Basic Wind Speed

In Eq. 6.1 the wind climate is explicitly represented by the basic wind speed v_{50}. Different v_{50} values are specified to account for regional differences in climate. In SANS 10160-3:2019, one of four values $v_{50} \in \{32, 36, 40, 44\}$ m/s are specified for each district and metropolitan municipality (administrative divisions of a South African province) using a characteristic wind speed map (Figure 6.1) and table. SANS 10160-3:2019 introduced this map, which represents a substantial improvement over previous maps. Before Figure 6.1

was adopted, the characteristic map was based on studies which only had access to sparse data (Milford, 1985*b*, 1987*b*).

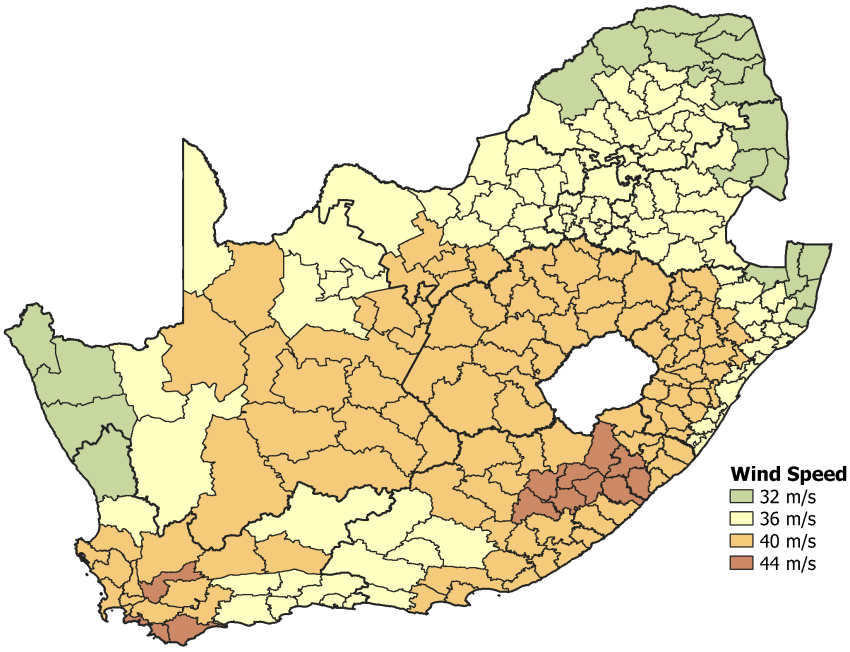


Figure 6.1: Characteristic wind speed map in SANS 10160-3:2019.

The current map originates from an analysis of gust wind speeds collected at 92 weather stations throughout South Africa performed by Kruger (2011), who estimated the annual $p = 0.02$ exceedance gust wind speeds using Peak-Over-Threshold (POT) and Gumbel analyses (Palutikof *et al.*, 1999).

Cook (1985) recommends at least 20 years of data for reliable Gumbel results, but data used by Kruger (2011) was limited to the period between 1990 and 2008, with a longest series of only 16 years, which is why Kruger (2011) also performed the POT analysis. To be conservative Kruger (2011) then adopted the maximum 75th percentile gust wind speed of the Gumbel and POT analysis.

In response to the shortage of data, an effort to increase the quantity of available quality data has been made (Chapter 3). This yielded an extended dataset which includes over 3500 annual gust maxima from 132 stations classified by dominant storm type and normalised to standardised conditions. This is more than a twofold increase in data

compared to Kruger (2011)'s study. The location and dominating extreme wind speed mechanism of each station is shown in Figure 6.2.

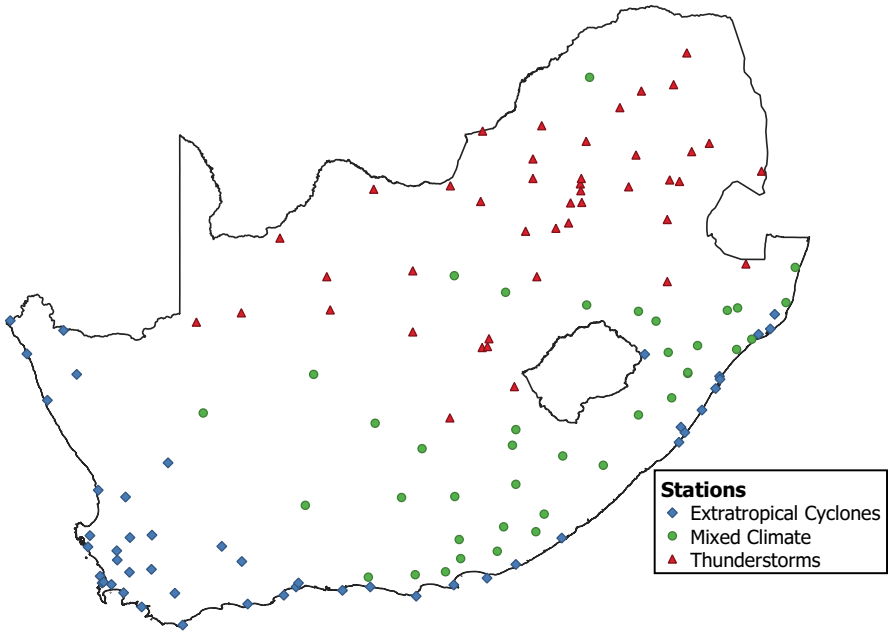


Figure 6.2: Distribution and dominant climatic mechanism of stations used in this study.

Chapter 4 considered the extended dataset from Chapter 3 in terms of model selection and found that a Gumbel analysis was preferred. The downside of using the Gumbel distribution is that its inflexibility relative to distributions with more parameters, such as the Generalised Extreme Value distribution, could introduce significant model-bias. To reduce model-bias without a significant increase in model-variance, Chapter 4 recommended regional *preconditioning* of the wind speed (raising the wind speeds by an exponent) before fitting the Gumbel distribution. It was found that fitting the Gumbel distribution to wind speeds raised by an exponent of 1.6 should lead to a reduction of model-bias, given the extended South African dataset. This technique and the new data could be used to update the v_{50} values in SANS 10160-3, which should improve its representation of the extreme wind climate.

6.1.2 Reliability Performance

In Eq 6.1 the wind climate is also implicitly represented by the partial factor γ_w. The γ_w accounts for uncertainties in the load, a significant portion of which can be attributed to the wind climate (Ellingwood and Tekie, 1999; Holickỳ, 2009; Botha *et al.*, 2018*b*). A reliability assessment by Botha *et al.* (2018*a*) led to the $\gamma_w = 1.6$ currently used in SANS 10160-1:2019.

A reliability assessment aims to ensure that a structure designed according to the standard has an acceptably low lifetime probability of failure. This requirement can be considered for a wind loading standard using a limit state l of

$$\begin{aligned} l &= r - g - q, \\ q &= cv^2. \end{aligned} \tag{6.2}$$

Where r, g and q are random variables representing the structural resistance, permanent loading and wind loading respectively. The wind loading can be decomposed into two separate random variables: the wind load modification factor c and the wind speed v.

An acceptable design is found by specifying a set of design values $\{r_d, g_d, q_d\}$ which ensure that the probability of a load exceeding the resistance $P(l < 0)$ equals some target probability p_t. This p_t is associated with a specified target level of reliability and is typically referred to using the target reliability index $\beta_t = \Phi^{-1}(1 - p_t)$, with Φ the standard cumulative normal distribution function. Because structures differ in performance, consequence of failure, and nature of failure, different β_t values are specified for different reliability classes (Retief and Dunaiski, 2009). Deviation from these β_t values should be minimised throughout the scope of the standard.

If distributions of the random variables can be defined, then a set of optimal design values, at the most probable point of failure, can be found. An efficient way to achieve this is provided by the First Order Reliability Method (FORM) (Ang and Tang, 1984).

Botha *et al.* (2018*a*) considered a single representative distribution of dynamic wind pressure (equivalent to v^2) and a range of g, r, and c distributions in their reliability

assessment of SANS 10160-3. Unlike the probability distribution of r, g and c, which can reasonably be considered location independent, the distribution of v could differ by climatic region and thus vary by location (Ellingwood and Tekie, 1999). This implies that using a single distribution of v to derive a single γ_w may lead to geographical variation in reliability performance. This may be acceptable if the region covered by the standard has a fairly homogeneous wind climate (Hong *et al.*, 2016), as is the case for many countries that use EN 1991-1-4.

Larger regions are likely to experience substantial variation in the wind climate, and could see significant geographical variation in reliability performance if a single γ_w is specified for the entire region. Therefore, standards that aim to cover diverse climatic regions, including the American (ASCE 7-16) and Australian (AS/NZS 1170.2) standard, adapted the partial factor based format of Eq. 6.1.

Instead, a high return period design wind speed v_d that corresponds with the target reliability is specified, effectively setting $\gamma_w = 1$. This ensures that as long as regionally representative statistics are used to estimate v_d, the regional variation in the wind climate is accounted for in the design (Hong *et al.*, 2016). The strong wind climate throughout South Africa is diverse (Kruger *et al.*, 2010), and so an approach similar to ASCE 7-16 and AS/NZS 1170.2 could be considered to yield more uniform reliability performance.

6.2 Estimation of Design Wind Speeds

6.2.1 Direct Consideration of Design Wind Speeds

The use of a single partial factor across an area with some regions dominated by synoptic storms and others by tropical cyclones is inappropriate because the variability of wind speeds from tropical cyclones is typically higher than synoptic storms. This has been recognised in standards that need to cover both cyclone and synoptic storm regions; in the United States and Australia an 'importance factor'/'cyclone factor' was used to to increase design loads to account for the higher variability of tropical cyclones (Holmes, 2018). In both standards this practice has been superseded and a design wind speed v_d

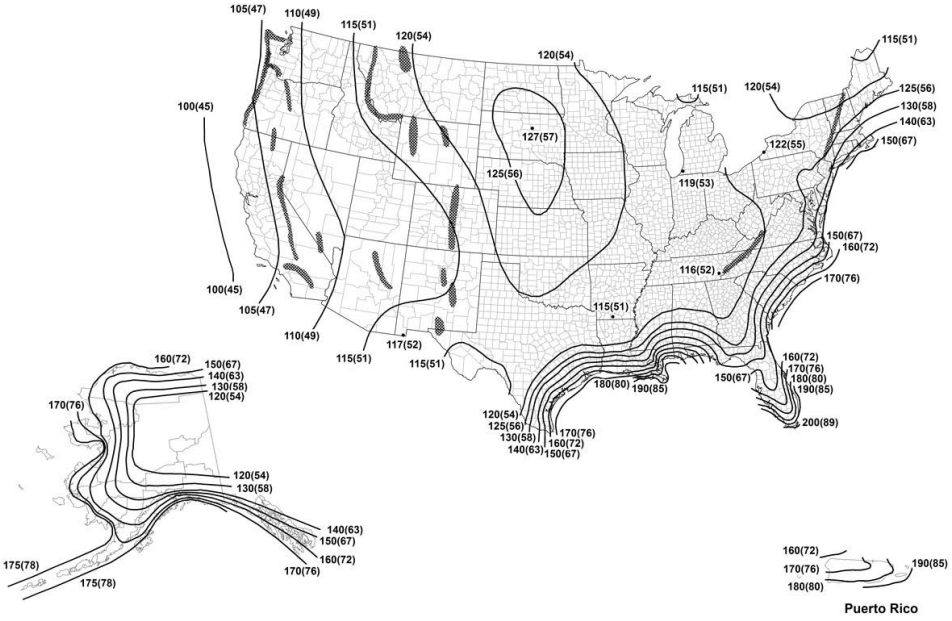


Figure 6.3: Design wind speed map in ASCE 7-16 for risk category 2 structures (700-year return period). Values are given in miles/hour and (metres/second).

is used instead. The exceedance probability of this design wind speed $p_{t:v_d}$ is specified to correspond to the target reliability. This avoids the need to have different regional partial factors because the regional differences are reflected in v_d.

The map of v_d values for the reference level of reliability in ASCE 7-16 is shown in Figure 6.3, here the differences in wind speeds between the east coast and the rest of the country are the result of the cyclonic region on the east coast. ASCE 7-16 and AS/NZS 1170.2 specify multiple v_d values at a given location depending on the reliability class (risk categories in ASCE 7-16 or importance levels in AS/NZS 1170.2) because the difference in variability between cyclonic and non-cyclonic winds is too large to reasonably allow conversion between $p_{t:v_d}$ values using a single formula.

South Africa has a diverse wind climate, dominated by synoptic winds in the south-west and thunderstorm winds in the north-east (Kruger *et al.*, 2010). These different climatic mechanisms were found to have significantly (P-value $\lessapprox 0.01$) different coefficients of variation of v^2 with means of $\overline{\delta}_{v^2:syn} = 0.24$ and $\overline{\delta}_{v^2:ts} = 0.29$ for synoptic and thunderstorm regions respectively (Botha, 2016).

This trend was further investigated using the dataset compiled in Chapter 3. The

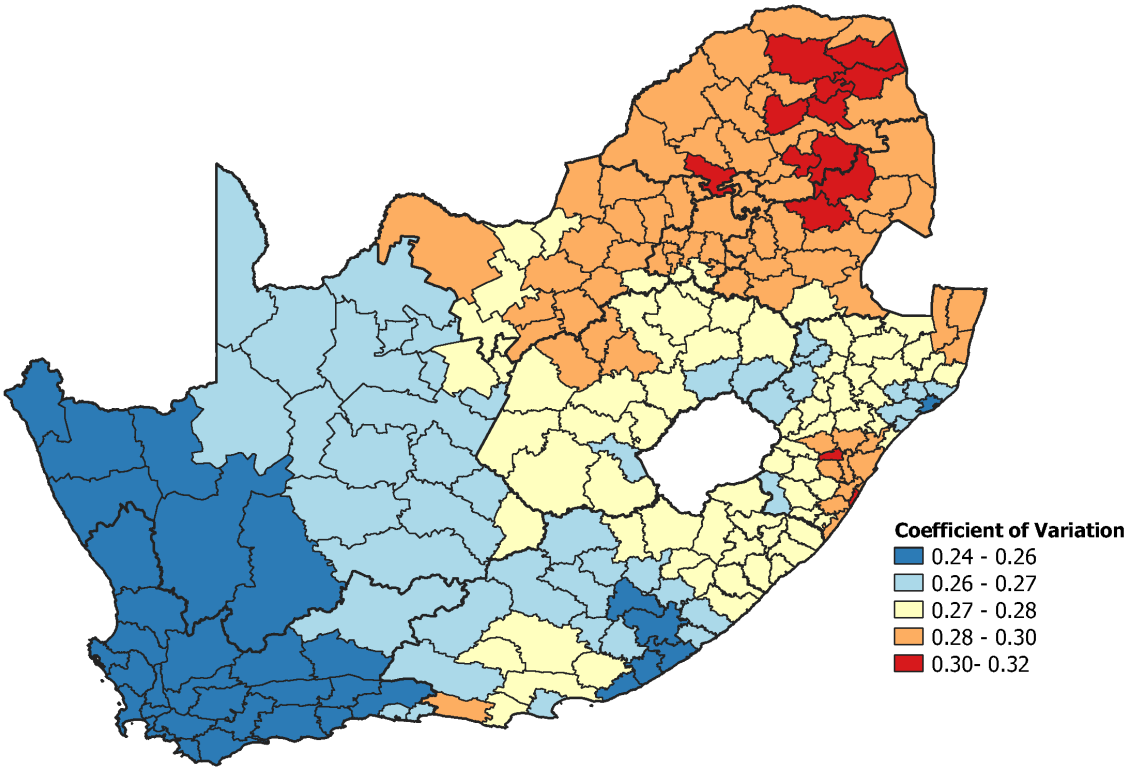


Figure 6.4: Interpolated coefficient of variation of v^2 throughout South Africa.

coefficient of variation δ_{v^2} of the 170 available series was estimated, then inverse distance weighting was used to interpolate between stations and δ_{v^2} was aggregated by municipality. This yielded a map of δ_{v^2}, shown in Figure 6.4. There is a clear increase in δ_{v^2} from south-west to north-east that correlates with the dominant climatic mechanism (Figure 6.2), confirming the trend observed by Botha (2016) of synoptic dominated regions having a lower δ_{v^2} and thunderstorm dominated regions a higher δ_{v^2}.

Botha *et al.* (2018*a*) aimed to perform a reliability assessment for the country as a whole and used $\delta_{v^2} = 0.31$ to calibrate the partial factor $\gamma_w = 1.6$ for the entire country. While this δ_{v^2} was a reasonable choice for nationwide consideration, Figure 6.4 suggests it may have been overly-conservative for synoptic regions in the south-west of the country. Therefore, direct consideration of v_d values is proposed to obtain more regionally representative design loads. An advantage of directly considering v_d instead of the characteristic values ($p = 0.02$) is that statistical techniques can be applied directly and the full impact of the new information (Chapter 3, Chapter 4, Chapter 5) can be

realised.

The v_d should be estimated for an exceedance probability $p_{t:v_d}$ that will result in attainment of the target reliability. This $p_{t:v_d}$ could be found by performing a reliability assessment using FORM and finding the $p_{t:v_d}$ from the design point. SANS 10160-3:2019 is meant to apply to a wide range of design situations (Retief and Dunaiski, 2009) and so multiple resistance distributions and wind to total load ratio's would need to be considered, which would lead to ambiguity about which design point to use.

Alternatively, because Botha *et al.* (2018*a*) has already calibrated the reliability level of the standard, the $p_{t:v_d}$ currently provided by SANS 10160-3:2019 could be calculated and used instead. This would avoid having to repeat the subjective decisions and calibration procedure done by Botha *et al.* (2018*a*).

6.2.2 ASCE Exceedance Probability

The $p_{t:v_d}$ used by ASCE 7-16 was back-calculated from provisions in ASCE 7-05 which used a similar format to SANS 10160-3:2019 (Cook *et al.*, 2011) with design wind loads also calculated using Eq. 6.1. If it is assumed that the variability of the wind speed dominates the uncertainties accounted for by the partial factor γ_w, an alternative formulation using a design wind speed v_d is

$$q_d = c_k v_d^2. \tag{6.3}$$

Combining this with Eq. 6.1 means v_d can be represented in terms of the previous provisions as

$$v_d = \sqrt{\gamma_w} v_{50}. \tag{6.4}$$

ASCE 7-05 allowed adjustment of the characteristic value to another annual exceedance probability p using

$$v_k = c_{prob} v_{50}, \tag{6.5}$$

where

$$c_{prob:asce} = \left[0.36 + 0.1 \ln(12/p)\right]. \tag{6.6}$$

Combing this with Eq. (6.4) gives

$$\sqrt{\gamma_w} = [0.36 + 0.1 \ln(12/p_{t:v_d})] \,, \tag{6.7}$$

from which the $p_{t:v_d}$ that matches the reliability of the wind speeds can be determined. For example, the ASCE 7-16 reference level of reliability (Risk Category II) with $\gamma_w = 1.6$ gives $p_{t:v_d} = 1.4 \times 10^{-3}$ (700-year return period). The similarities between ASCE 7-05 and SANS 10160-3:2019 means that the same procedure can be considered for South Africa.

In SANS 10160-3:2019 the adjustment of v_{50} to a different annual exceedance probability is done with

$$c_{prob:sans} = \left[\frac{1 - 0.2\ln(-\ln(1-p))}{1 - 0.2\ln(-\ln(0.98))}\right]^{0.5} . \tag{6.8}$$

The relationship between p_t and γ_w resulting from c_{prob} is depicted in Figure 6.5, where a significant discrepancy between ASCE and SANS is evident. The partial factor in SANS 10160-1:2019 is also $\gamma_w = 1.6$, but $c_{prob:sans}$ implies an annual $p_{t:v_d} = 10^{-4}$, which is more conservative than ASCE by an order of magnitude.

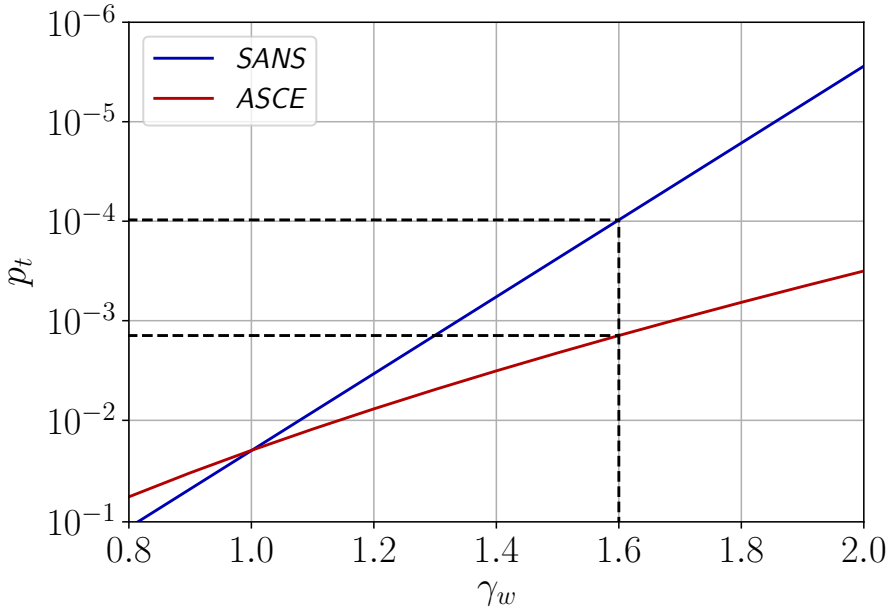


Figure 6.5: Relationship between exceedance probability and partial factor, using the method described by Cook *et al.* (2011).

This discrepancy led to an investigation into $c_{prob:sans}$, summarised in Appendix 6.B, which found that

$$c^*_{prob:sans} = \left[\frac{1 - 0.2\ln(-\ln(1-p))}{1 - 0.2\ln(-\ln(0.98))}\right]^{0.63}, \tag{6.9}$$

could be more appropriate given typical South African conditions. This would increase the calculated $p_{t:v_d}$ in SANS to $p_{t:v_d} = 3.6 \times 10^{-4}$, which is still significantly different to ASCE 7-16.

The ASCE method of determining $p_{t:v_d}$ assumes that v dominates the uncertainty of the wind loading formulation. Prescribing all the uncertainty to v means that the variability of v dictates the design load q_d. Therefore, if uncertainties in c are large and the $p_{t:v_d}$ is calculated neglecting these uncertainties, then the $p_{t:v_d}$ obtained would be too far into the tail of the distribution resulting in too much weight being given to variability of v. Thus, q_d would be overestimated in areas where the variation of v is high and underestimated in areas where the variation of v is low.

Research by Botha *et al.* (2018*b*) found that uncertainties in other components of Davenport (1982)'s wind loading chain could be significant. These findings were incorporated into Botha *et al.* (2018*a*) calibration of SANS 10160-3; hence, neglecting uncertainties in c when back-calculating the target reliability is not considered reasonable.

6.2.3 Most Probable Exceedance Probability

If uncertainties in c are considered the design wind load can be represented as

$$q_d = c_d v_d^2, \tag{6.10}$$

which can be combined with Eq. (6.1) to give

$$\gamma_w = \left(\frac{c_d}{c_k}\right)\left(\frac{v_d^2}{v_{50}^2}\right). \tag{6.11}$$

Unlike the ASCE method there are multiple possible v_d values, with corresponding c_d values, which together satisfy Eq. 6.11.

Defining exceedance of the design load as failure means that

$$l = \gamma_w - \left(\frac{c}{c_k}\right)\left(\frac{v^2}{v_{50}^2}\right) \tag{6.12}$$

represents an acceptable limit state, with c and v as random variables. Assuming $\{v_d, c_d\}$ are at the design point, solving $P(l < 0)$ using FORM provides both the probability β_q that the wind load exceeds its design value and the most probable $\{v_d, c_d\}$ solution. The exceedance probabilities of c_d and v_d are then defined by their sensitivity factors, for example $p_{t:v_d} = 1 - \Phi(\alpha_v \beta_q)$.

The limit state can be formulated in terms of the regionally preconditioned wind speeds $v^{\overline{w}}$ as

$$l = \gamma_w - \left(\frac{c}{c_k}\right)\left(\frac{v^{\overline{w}}}{v_{50}^{\overline{w}}}\right)^{2/\overline{w}}. \tag{6.13}$$

Setting $\overline{w}$ to the most likely value estimated in Section 4.3.2 of $\overline{w} \approx 1.6$ is justified by the small impact that $\overline{w}$ values within the 90% confidence interval were shown to have on the design values in Section 4.4. Then since in South Africa $\gamma_w = 1.6$

$$l = 1.6 - \left(\frac{c}{c_k}\right)\left(\frac{v^{1.6}}{v_{50}^{1.6}}\right)^{2/1.6}. \tag{6.14}$$

The FORM $P(l < 0)$ solution can then be found if the distributions of c and $v^{\overline{w}}$ relative to their values in the standard c/c_k and $v^{1.6}/v_{50}^{1.6}$ are defined.

6.2.3.1 Distribution of v

A generalisation of the Gumbel distribution (2.3.3) to $v^{\overline{w}}$ has a cumulative density function of

$$1 - p = \exp\left[-\exp\left\{-\frac{(\pi v^{\overline{w}} - \pi\mu_{v^{\overline{w}}} + \gamma\sqrt{6}\sigma_{v^{\overline{w}}})}{\sqrt{6}\sigma_{v^{\overline{w}}}}\right\}\right]. \tag{6.15}$$

where $\gamma = 0.577\ldots$ is Euler's gamma. The $\mu_{v^{\overline{w}}}$ and $\sigma_{v^{\overline{w}}}$ are the mean and standard deviation of the regionally preconditioned wind speed $v^{\overline{w}}$. Equation 6.15 can also be represented as

$$v = [\mu_{v^{\overline{w}}} + \theta(p)\sigma_{v^{\overline{w}}}]^{1/\overline{w}}, \tag{6.16}$$

where θ is a standardisation function given as

$$\theta(p) = -\frac{\sqrt{6}}{\pi}\left[\gamma + \ln(-\ln(1-p))\right]. \tag{6.17}$$

The $v^{\overline{w}}/v_{50}^{\overline{w}}$ would still be Gumbel distributed, with mean

$$\begin{aligned} \mathrm{E}[v^{\overline{w}}/v_{50}^{\overline{w}}] &= \frac{\mu_{v^{\overline{w}}}}{\mu_{v^{\overline{w}}} + \theta(0.02)\sigma_{v^{\overline{w}}}} \\ &= \frac{1}{1 + 2.59\delta_{v^{\overline{w}}}}, \end{aligned} \tag{6.18}$$

and standard deviation

$$\begin{aligned} \mathrm{STD}[v^{\overline{w}}/v_{50}^{\overline{w}}] &= \frac{\sigma_{v^{\overline{w}}}}{\mu_{v^{\overline{w}}} + \theta(0.02)\sigma_{v^{\overline{w}}}} \\ &= \frac{\delta_{v^{\overline{w}}}}{1 + 2.59\delta_{v^{\overline{w}}}}, \end{aligned} \tag{6.19}$$

where $\delta_{v^{\overline{w}}} = \sigma_{v^{\overline{w}}}/\mu_{v^{\overline{w}}}$ is the coefficient of variation of $v^{\overline{w}}$.

Botha *et al.* (2018*a*) used a $\delta_{v^2} = 0.31$, to represent the variation of v^2. This annual coefficient of variation is equivalent to $\delta_{v^{1.6}} = 0.25$. The distribution of $v^{1.6}/v_{50}^{1.6}$ can therefore be reasonably represented by a Gumbel distribution with mean $\mathrm{E}(v^{1.6}/v_{50}^{1.6}) = 0.6$ and standard deviation $\mathrm{STD}(v^{1.6}/v_{50}^{1.6}) = 0.15$ defined by Eqs. 6.18 and 6.19.

6.2.3.2 Distribution of c

Botha *et al.* (2018*a*) specified each c/c_k component c_i/c_{ki} as having a normal distribution, with mean $\mathrm{E}[c_i/c_{ki}]$ and standard deviation $\mathrm{STD}(c_i/c_{ki})$. Botha *et al.* (2018*a*) was mainly concerned with estimating $\mathrm{E}[c_i/c_{ki}]$ and $\mathrm{STD}(c_i/c_{ki})$ rather than the specific distribution of c/c_k.

A log-normal distribution has also been used for c_i/c_{ki} (JCSS Model Code; Baravalle and Köhler, 2018; Hong *et al.*, 2016). The product of a series of log-normally distributed random variables also follows a log-normal distribution, with parameters defined by it's component distributions. Therefore, if each c_i/c_{ki} is assumed to be log-normally distributed with mean and standard deviation defined by Botha (2016), then these can

be combined (Castillo *et al.*, 2005) to find the distribution of c/c_k.

Botha *et al.* (2018*a*) considered four different sets of c_i/c_{ki} distributions, resulting in four probabilistic models. The first model (New-SANS) was the result of a detailed study into wind load uncertainties by Botha (2016), who found that uncertainties in c/c_k could be significantly higher than typically assumed.

Botha *et al.* (2018*a*) was concerned that, given the significant increase in variability of c/c_k, using only the New-SANS model in a reliability assessment may be too drastic. Instead a Bayesian philosophy was employed, where the New-SANS model was taken as a prior and used to update three other existing c/c_k models by Holickỳ (2009), Milford (1985*b*), and Gulvanessian and Holicky (2005). The c_i/c_{ki} components of each of Botha *et al.* (2018*a*)'s four models (New-SANS and the three updated models) were assumed to be log-normally distributed and combined to give four log-normal c/c_k distributions, with means $\mathrm{E}(c/c_k)$ and standard deviations $\mathrm{STD}(c/c_k)$ given in Table 6.1.

6.2.3.3 FORM Results

Using the limit state function (Eq. 6.14) and defined $v^{1.6}/v_{50}^{1.6}$ and c/c_k distributions, a FORM analysis was performed. The results for each of Botha *et al.* (2018*a*)'s probabilistic models are given in Table 6.1.

Table 6.1: Mean and standard deviation of combined lognormal c/c_k distributions defined by Botha *et al.* (2018*a*) and results of FORM analysis.

Model	$\mathrm{E}(c/c_k)$	$\mathrm{STD}(c/c_k)$	β_q	α_v	α_c	c_d/c_k	$p_{t:v_d}$
New-SANS	0.7	0.28	3.11	0.68	0.73	1.56	16.7×10^{-3}
Updated-Milford	0.57	0.18	3.83	0.77	0.64	1.16	1.7×10^{-3}
Updated-Gulv/Holický	0.67	0.19	3.61	0.80	0.60	1.18	2.0×10^{-3}
Updated-Holický	0.84	0.19	3.32	0.85	0.52	1.2	2.3×10^{-3}

For all models these results imply that the wind load dominates the uncertainty of the overall formulation, with the annual exceedance probability of the design wind load β_q between 78 % and 96 % of the annual reference target exceedance probability $\beta_t = 4.0$. The high uncertainty of c/c_k in the New-SANS models means that, unlike the other three

models, $\alpha_c > \alpha_v$ indicating that c contributed more to the total uncertainty of the wind load than v. This directly contradicts the ASCE method, where $\alpha_c = 0$ was assumed.

The updated models all have a fairly consistent design point with an effective partial factor for other components of the design load $c_d/c_k \approx 1.2$ and design wind speed annual exceedance probability $p_{t:v_d} \approx 2 \times 10^{-3}$ (500-year return period). Given that Botha *et al.* (2018*a*) did not advocate the direct use of the New-SANS model, the design point indicated by the updated-models is accepted as representative of the overall reliability currently offered by SANS 10160-3:2019.

6.2.4 Optimal Estimation Procedure

If only the recordings from a particular site are used to estimate v_d, then more representative values are obtained on average (low bias), although this introduces substantial model-variance (Hong *et al.*, 2016). A lower model-variance alternative is to regionally aggregate the data (Hosking and Wallis, 2005), although this risks higher model-bias.

Chapter 5 considered this problem in terms of the bias variance trade-off (Friedman *et al.*, 2001). It was found that by employing an optimal combination of site and regional statistics, a design value which better corresponds to $p_{t:v_d}$ could be found. This optimal estimate $\hat{v}_d$ is found using sample estimates of mean $\hat{\mu}_{v^{1.6}}$ and standard deviation $\hat{\sigma}_{v^{1.6}}$ from the n regionally preconditioned annual maximum wind speeds measured at a particular site, located in a region with an average coefficient of variation $\overline{\delta}_{v^{1.6}}$, from Eq. 5.37, as

$$\hat{v}_d = \left[\frac{1 + \theta_t \overline{\delta}_{v^{1.6}}}{1 + \theta_{k:opt} \overline{\delta}_{v^{1.6}}} (\hat{\mu}_{v^{1.6}} + \theta_{k:opt} \hat{\sigma}_{v^{1.6}})\right]^{1/1.6}, \tag{6.20}$$

where from Eq. 5.34

$$\theta_{k:opt} = \frac{\delta^2_{v^{1.6}} \left(1 + \theta_t \overline{\delta}_{v^{1.6}}\right) \left(2\lambda - 4\overline{\delta}_{v^{1.6}}\right) - 4(\overline{\delta}_{v^{1.6}} - \delta_{v^{1.6}})^2 n \theta_t}{\delta^2_{v^{1.6}} \left(1 + \theta_t \overline{\delta}_{v^{1.6}}\right) \left(2\overline{\delta}_{v^{1.6}} \lambda - \kappa + 1\right) - 4(\overline{\delta}_{v^{1.6}} - \delta_{v^{1.6}})^2 n}. \tag{6.21}$$

In this formulation $\theta_t = \theta(p_{t:v_d})$ is the standard score, i.e. the number of standard deviations away from the mean, of the design value and the λ and κ are the skewness and

kurtosis of the Gumbel distribution, equal to 1.14 and 5.4 respectively.

The $\overline{\delta}_{v^{1.6}}$ can be estimated by weighting the estimated coefficient of variation of each of the N series $\hat{\delta}_{v^{1.6}i}$ by its sample size n_i as suggested by Hosking and Wallis (2005)

$$\hat{\overline{\delta}}_{v^{1.6}} = \frac{\sum_{i=1}^{N} n_i \hat{\delta}_{v^{1.6}i}}{\sum_{i=1}^{N} n_i}, \tag{6.22}$$

and if N is large enough ($\gtrapprox 20$) then $\overline{\delta}_{v^{1.6}} \approx \hat{\overline{\delta}}_{v^{1.6}}$. In Chapter 5 it was recommended that thunderstorm and synoptic dominated areas be considered as separate regions with their own $\overline{\delta}_{v^{1.6}}$. The 83 thunderstorm series yield $\overline{\delta}_{v^{1.6}} = 0.25$ and the 87 synoptic series yield $\overline{\delta}_{v^{1.6}} = 0.18$.

An estimate of the design wind speed $\hat{v}_d$ can thus be found for a $p_{t:v_d}$. This can be expressed as $\hat{v}_d = f(p_{t:v_d}, \Theta)$ where f denotes that $\hat{v}_d$ is a function of $p_{t:v_d}$ and some other parameters Θ dependant on the available data. In sites with a single mechanism $\Theta = \{\hat{\mu}_{v^{1.6}}, \hat{\sigma}_{v^{1.6}}, n, \overline{\delta}_{v^{1.6}}\}$ and in sites with two mechanisms the parameters from both need to be considered $\Theta = \{\Theta_1, \Theta_2\}$. The $\hat{v}_d$ at sites with more than one mechanism can be found by accounting for multiple mechanisms, as recommended by Gomes and Vickery (1978), described in Section 2.3.2.

The procedure outlined in this Section provides a solution for the estimation of v_d at weather stations throughout South Africa, but it is still unclear how to extend this information for the entire country and incorporate it into the standard.

6.3 Updating Design Wind Loads in SANS

The current wind loads in SANS 10160-3:2019 are the culmination of a significant research effort that includes detailed surface roughness correction (Kruger *et al.*, 2011), climatic classification (Kruger *et al.*, 2010), statistical modelling (Kruger *et al.*, 2013*a*), mapping considerations (Kruger *et al.*, 2013*b*, 2017), and expert engineering and climatological knowledge.

The estimation procedure outlined in Section 6.2 with some interpolation procedure

(Ye *et al.*, 2015) could be used to replace the current map, although much of the information contained in it would be lost. Alternatively, the current map could be used as a basis from which changes are recommended where new data (Chapter 3) and statistical techniques (Chapter 4, Chapter 5) provide a strong enough indication that the current values are inappropriate. This approach will incorporate the new information and provide a solution to the extension of the design values throughout the country, while maximising the amount of expert knowledge retained.

6.3.1 Comparison of Estimated Design Loads to SANS

Using the design point found in Section 6.2.3 and the statistics from the available data Θ, v_d can be estimated at each station using the optimal estimation procedure described in Section 6.2.4 $\hat{v}_d = f(2 \times 10^{-3}, \Theta)$. An estimated design load $\hat{q}_d$ for each station can then be obtained using Eqs. 6.1 and 6.11

$$\begin{aligned} \hat{q}_d &= \frac{1}{2}\rho \left(\frac{c_d}{c_k}\right) \left(\frac{\hat{v}_d^2}{v_{50}^2}\right) c_k v_{50}^2 \\ &= \frac{1}{2}\rho 1.2 c_k \hat{v}_d^2. \end{aligned} \tag{6.23}$$

This can be compared to the design wind load currently specified by SANS 10160-3:2019

$$q_{d:sans} = \frac{1}{2}\rho 1.6 c_k v_{50}^2, \tag{6.24}$$

where v_{50} is found using the characteristic wind map in SANS 10160-3:2019 (Figure 6.1). The difference between the predicted $\hat{q}_d$ and the current $q_{d:sans}$ design load can be assessed as

$$\begin{aligned} \Delta_{sans} &= \frac{\hat{q}_d - q_{d:sans}}{q_{d:sans}} \\ &= \frac{1.2\hat{v}_d^2 - 1.6 v_{50}^2}{1.6 v_{50}^2}. \end{aligned} \tag{6.25}$$

The distribution of Δ_{sans} values from all the stations are summarised in Figure 6.6. These show that SANS 10160-3:2019 is conservative overall with the average $\Delta_{sans} = -10$ %,

which could be attributed to the conservative measures taken by Kruger (2011) or the use of wind speed intervals in the map (Botha *et al.*, 2018*b*). There are $\hat{q}_d$ values that differ significantly from $q_{d:sans}$, suggesting that some could be reconsidered.

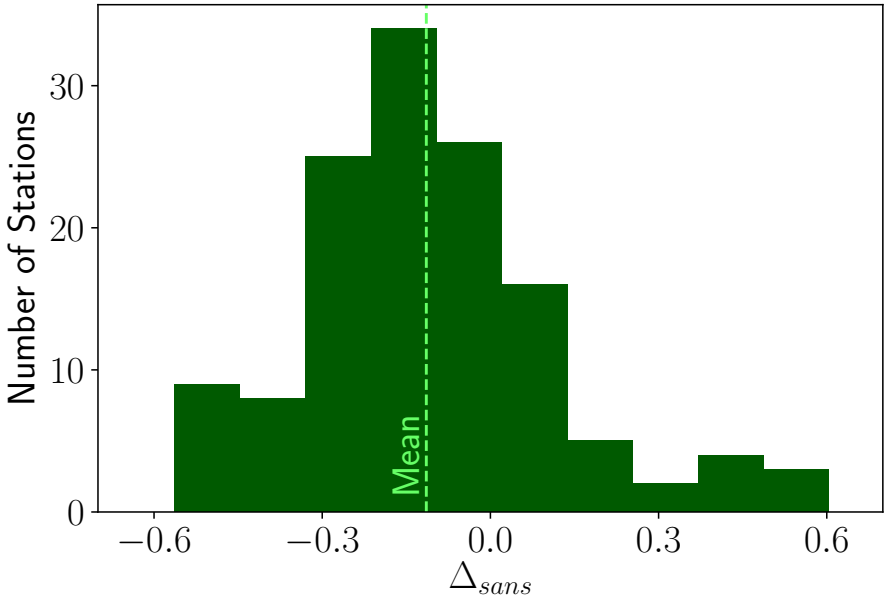


Figure 6.6: Difference in estimated design load and the load currently specified in SANS 10160-3:2019.

6.3.2 Identification of Stations with Unacceptable Design Wind Speeds

The accuracy of $\hat{v}_d$ is dependant on the quantity and quality of the v data used. Therefore, updating $q_{d:sans}$ to closely match $\hat{q}_d$ may not always result in the most appropriate design loads.

Large differences between $q_{d:sans}$ and $\hat{q}_d$ present evidence that $q_{d:sans}$ should be changed, but stations with more data present stronger evidence because there is higher confidence in these estimates. To determine at which stations there is enough evidence to reconsider $q_{d:sans}$ a statistical test was developed. This is based on the null hypothesis that the

current design value is correct

$$\mathrm{H}_0 : q_d - q_{d:sans} = 0. \tag{6.26}$$

If the design point found in Section 6.2.3 is accepted, then this H_0 is equivalent to

$$\mathrm{H}_0 : 1.2v_d^2 - 1.6v_{50}^2 = 0. \tag{6.27}$$

A statistical test with this null hypothesis means sufficient evidence must be presented before modification of the current standard is considered.

This test is founded on the assumption that $v^{1.6}$ is Gumbel distributed and that $\hat{\mu}_{v^{1.6}}$ and $\hat{\sigma}_{v^{1.6}}$ are fairly unbiased estimators of the mean and standard deviation (Chapter 4, Chapter 5). To test H_0, a set of simulated $\hat{v}_d$ values is obtained through Monte Carlo simulation:

1. For a series of n representative wind speed v observations available at a station, it is assumed that $v^{1.6}$ follows a Gumbel distribution with mean $\hat{\mu}_{v^{1.6}}$ and standard deviation $\hat{\sigma}_{v^{1.6}}$ equal to the sample estimates at the station.

2. Then n simulated wind speeds v_{sim} are drawn from this Gumbel distribution.

3. The mean $\hat{\mu}_{v^{1.6}:sim}$ and standard deviation $\hat{\sigma}_{v^{1.6}:sim}$ of v_{sim} are estimated.

4. The $\Theta_{sim} = \{\hat{\mu}_{v^{1.6}:sim}, \hat{\sigma}_{v^{1.6}:sim}, n, \overline{\delta}_{v^{1.6}}\}$ is thus available. For stations in a mixed climate, steps 1 to 3 are performed for each separate mechanism and $\Theta_{sim} = \{\Theta_{sim:1}, \Theta_{sim:2}\}$.

5. A simulated $\hat{v}_d$ is then $\hat{v}_{d:sim} = f(p_{t:v_d}, \Theta_{sim})$ (Section 6.2.4).

6. Steps 2 to 5 are repeated m times to obtain a set of m simulated $\hat{v}_{d:sim}$ values.

The proportion r of simulated design values which would result in a load less than $q_{d:sans}$ is then calculated as

$$r = \frac{1}{m}\sum_{i=1}^{m} \begin{cases} 1: & \text{if}(1.2\hat{v}_{d:sim}^2 < 1.6v_{50}^2) \\ 0: & \text{otherwise.} \end{cases} \tag{6.28}$$

If m is large enough then a P-value can be found

$$P\text{-value} = 1 - |2r - 1|. \tag{6.29}$$

A significance level $\boldsymbol{\alpha}^*$ can be selected and if P-value $< \boldsymbol{\alpha}^*$, then H_0 can be rejected, and a change of $q_{d:sans}$ at that station considered. The r value also indicates whether an increase $r < \boldsymbol{\alpha}^*/2$ or decrease $r > 1 - \boldsymbol{\alpha}^*/2$ in $q_{d:sans}$ should be considered.

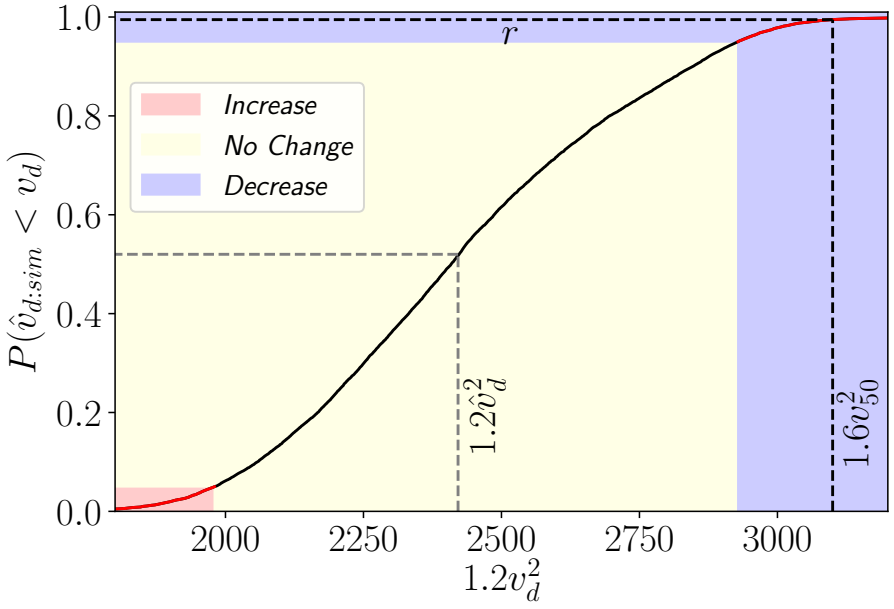


Figure 6.7: Empirical cumulative distribution of simulated design values at the Worcester station, m =10 000.

To demonstrate application of the test, an example showing how r, $\hat{v}_d$, and v_{50} relate to the empirical distribution of $\hat{v}_{d:sim}$ at the Worcester station is shown in Figure 6.7, where m =10 000 was used. An r value of 0.995 was obtained which means that the current design value can be rejected at a high significance level and that a decrease in the prescribed wind speed could be considered.

6.3.3 Recommended Changes to SANS 10160-3

Chapter 3 described the quality control of the data. This involved visual inspection of each gust to remove erroneous measurements, surface roughness correction, and accounting for the dynamic response of data recorded using a Dines anemometer. Chapter 3 did not account for topographical features or obstructions around instruments and so some stations with poor exposure cannot be considered representative of typical climatic conditions. Of the 132 stations, 102 have data with acceptable exposure.

Figure 6.8 overlays the results of the hypothesis test developed in Section 6.3.2 applied to each of these 102 stations onto the current characteristic map. On this figure, the shape and colour indicate whether an increase or decrease would be appropriate (red triangles vs blue inverted triangles) according to r, and the size of the symbol indicates the strength of the evidence against the current value (P-value). This figure will be used to guide recommended changes to SANS 10160-3:2019.

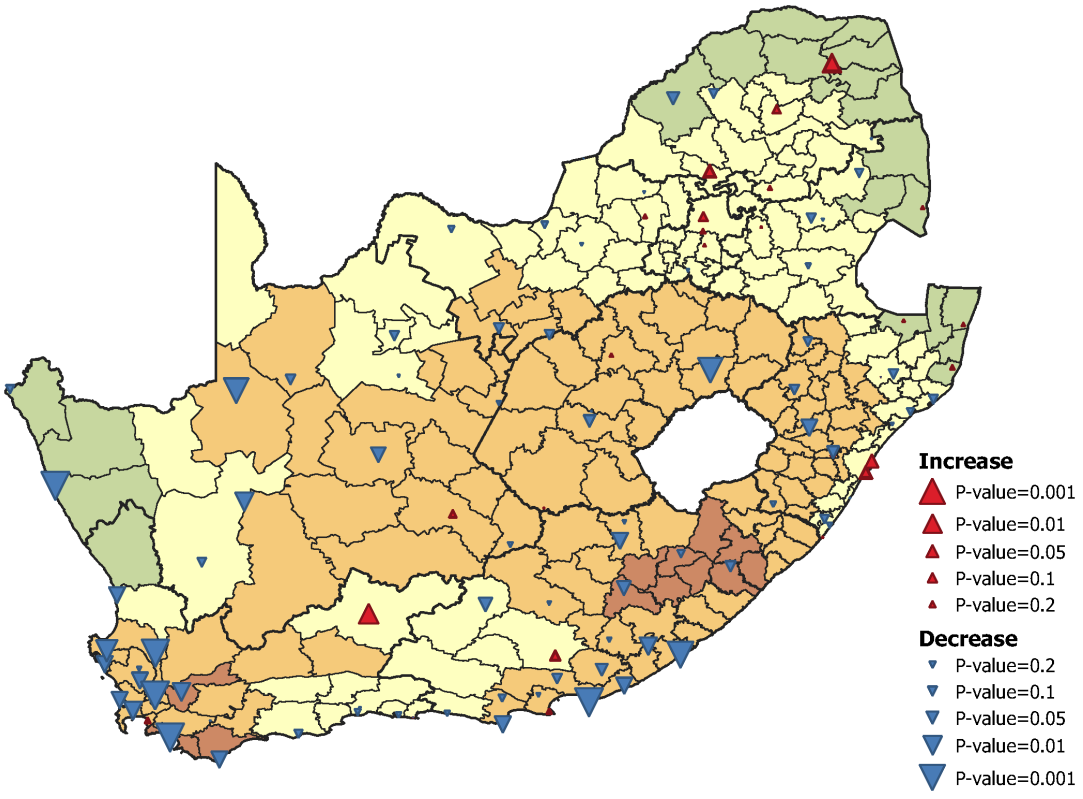


Figure 6.8: Evidence for increasing or decreasing the current wind speeds in SANS 10160-3:2019. Triangles are plotted at station locations.

Figure 6.8 shows that while there are some isolated municipalities where an increase could be considered, notably the Ethekwini (Durban) metropolitan municipality, there is strong evidence to support the reduction of $q_{d:sans}$ throughout much of South Africa. This is especially clear in the south-west of the country where synoptic winds dominate the extreme wind climate, such as the western parts of the Western Cape and a stretch along the Eastern Cape coast. The indication of a decrease in these areas can be attributed to the lower coefficient of variation in these areas, which means that $\gamma_w = 1.6$ was probably overly-conservative in these regions.

For 150 of the 234 municipalities there are no usable weather stations, and so no wind speed data is currently available. These municipalities tend to be rural with limited development, and thus the lack of data is not a major issue as SANS 10160-3:2019 aims to provide for the built environment.

To improve the standard, different $q_{d:sans}$ values which better represent the data could be recommended. The areas where change could be considered can be identified using Figure 6.8. To streamline the revision process and simplify implementation for practising engineers, it would be desirable if these changes can be made without significantly modifying the current format of the standard.

Specifying v_d values per reliability class as done in ASCE 7-16 and AS/NZS 1170.2 would require a substantial modification to the format of SANS 10160-3:2019. Synoptic and thunderstorm winds, which dominate South Africa, are more similar to each other than cyclonic winds (Holmes, 2018). Therefore, adjusting the design value from the reference reliability (annual $\beta_t = 4.0$) (Retief and Dunaiski, 2009) to other classes should not introduce significant error and only the v_d corresponding to $\beta_t = 4.0$ will be considered. This means that modification to the standard could be implemented by making adjustments to the characteristic wind speed map (Figure 6.1).

6.3.4 New Characteristic Wind Map

Implementing updates to the characteristic map will be a somewhat subjective process and require a degree of engineering judgement. As a starting point, $\boldsymbol{\alpha}^* = 0.1$ is used to

inform a first round of updates. The current v_{50} values are thus increased or decreased to an updated value $v_{50:updated}$ as

$$v_{50:updated} = \begin{cases} v_{50} + 4: & \text{if } r < 0.05 \\ v_{50} - 4: & \text{if } r > 0.95 \\ v_{50}: & \text{otherwise.} \end{cases} \tag{6.30}$$

It was found that 42 stations had a P-value < 0.1, indicating that at these locations there is reasonable justification to change the $q_{d:sans}$. For the majority of these stations a decrease in $q_{d:sans}$ could be considered, with $r > 0.95$ at 37 stations.

Some municipalities have multiple stations. In this case, a change was only implemented if it could be supported at all the stations in the municipality. This was the case for Buffalo City (East London), Ethekwini, and Saldanha Bay. If stations in a municipality contradicted each other, then the more conservative option was taken and only increasing v_{50} was considered. This was the case for the City of Cape Town and Msunduzi (Pietermaritzburg) where the current v_{50} were maintained despite one station indicating that a decrease could be appropriate.

The $v_{50:updated}$ values are shown in Figure 6.9 where municipalities with data are highlighted. Because the selection of $\boldsymbol{\alpha}^* = 0.1$ is somewhat arbitrary, this updated map is only meant to serve as a foundation for further changes.

The Koingnaas station located in Kamiesberg municipality indicates that a decrease could be considered at a 10% significance level, but here $v_{50} = 32$ m/s and could not be decreased without creating another v_{50} category. Since this is an isolated case, another category was not created and the current value was maintained $v_{50:updated} = 32$ m/s.

A noticeable difference between the updated map versus the current map is the reduction in municipalities with the highest $v_{50} = 44$ m/s wind speed category. There is only one station (Elliot) for which $v_{50} = 44$m/s cannot be rejected at a 10 % significance level. The data from this station still indicates that a decrease would not be unreasonable with $r = 0.92$. Therefore, without reasonable justification to maintain the $v_{50} = 44$ m/s category, it is recommended that it be eliminated.

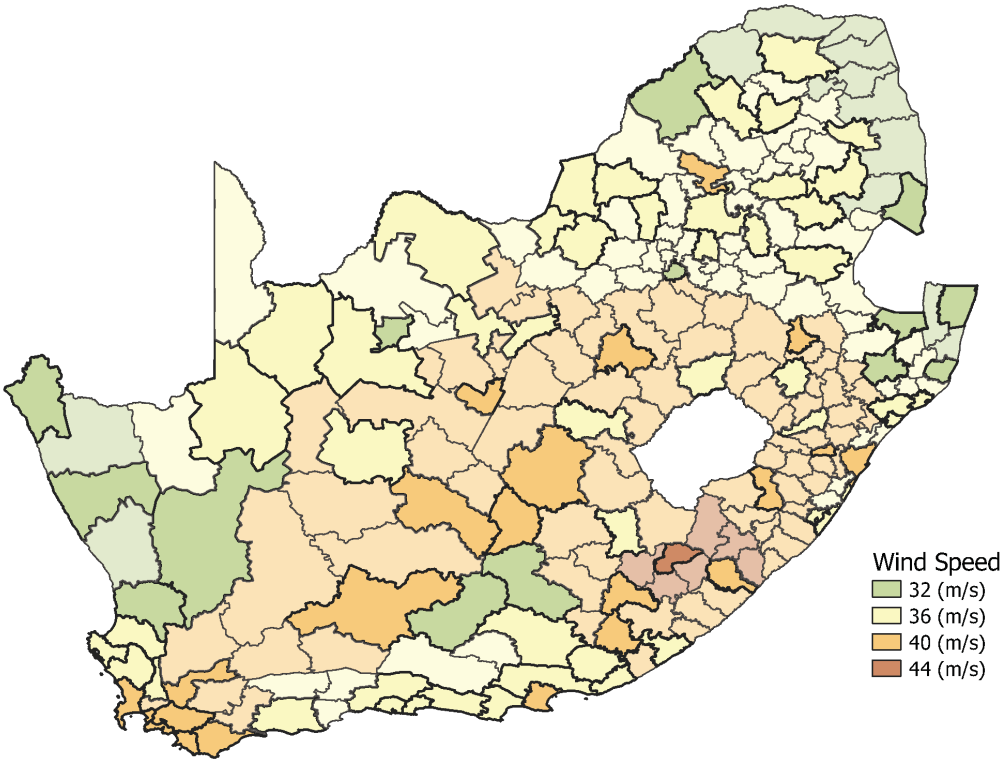


Figure 6.9: Updated characteristic map ($v_{50:updated}$). Municipalities with data are highlighted.

Historical data, recorded using a Dines anemometer, is available for a number of stations. This data is of lower quality due to uncertainties in the correction applied in Chapter 3. Inclusion of the Dines data results in less variation of $\hat{v}_{d:sim}$ values which leads to lower P-values. Given these uncertainties in the quality of this data, it alone was not considered sufficient evidence to modify $q_{d:sans}$. It was checked whether the more recent data, for which there is more certainty in the exposure and instrument response, could support a change independent of the Dines data. If it could not, as was the case at the Cape Town station, then changes were not made.

Kruger *et al.* (2017) assigned the current v_{50} values to municipalities without data using interpolation and engineering judgement. Therefore, it is argued that the v_{50} values at these municipalities can be changed to create a smoother map where necessary. There are also some isolated municipalities where the $v_{50:updated}$ value is surrounded by stations which generally support a different $v_{50:updated}$ value. These irregularities were removed by

adopting the more common $v_{50:updated}$ from surrounding stations.

The smoothing and elimination of $v_{50} = 44$ m/s in the updated map yielded a final recommended characteristic wind speed map, shown in Figure 6.10. The new map results in changes to 63 of the 234 municipalities throughout the country (Figure 6.11), of which 54 involve a decrease in v_{50}.

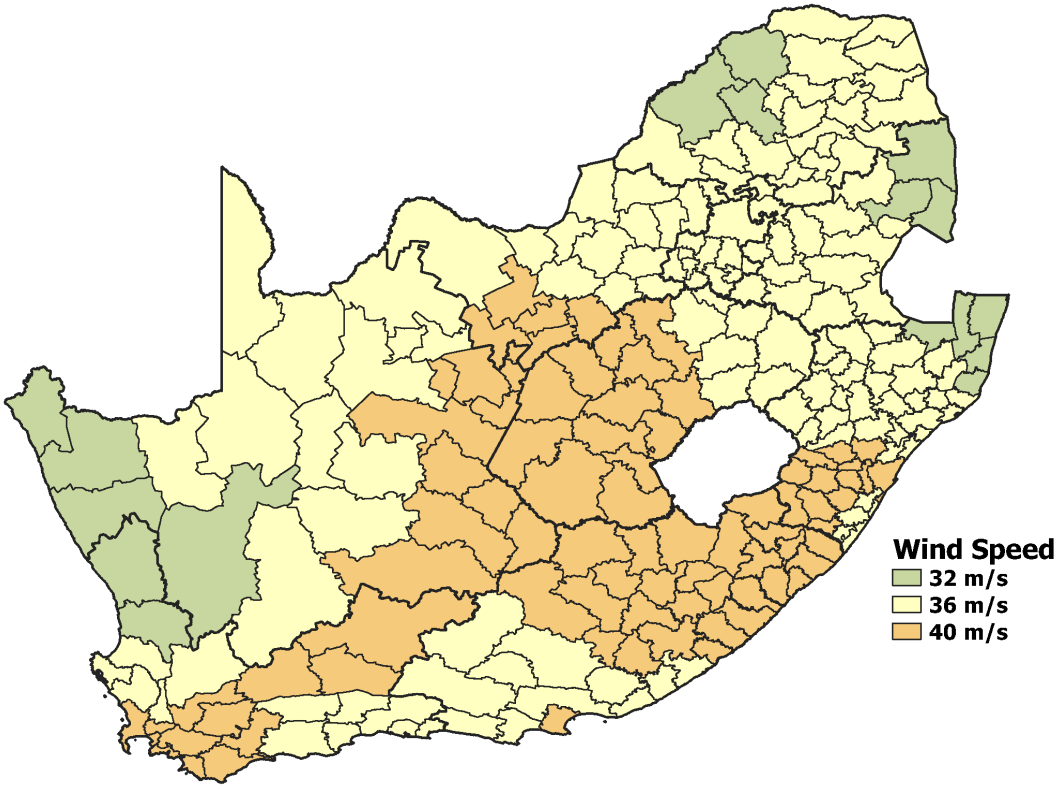


Figure 6.10: New characteristic wind speed map, recommended for future update to SANS 10160-3.

There are eight metropolitan municipalities in South Africa. These contain a large proportion of the population and built environment of the country which means changes to SANS 10160-3:2019 in these areas are of higher consequence. The new map recommends changes for the Ethekwini (Durban) and Buffalo City (East London) metropolitan municipalities. In both Ethekwini and Buffalo City a P-values < 0.1 was observed at two stations, which is considered to be sufficient evidence that the recommended changes for metropolitan municipalities could be adopted.

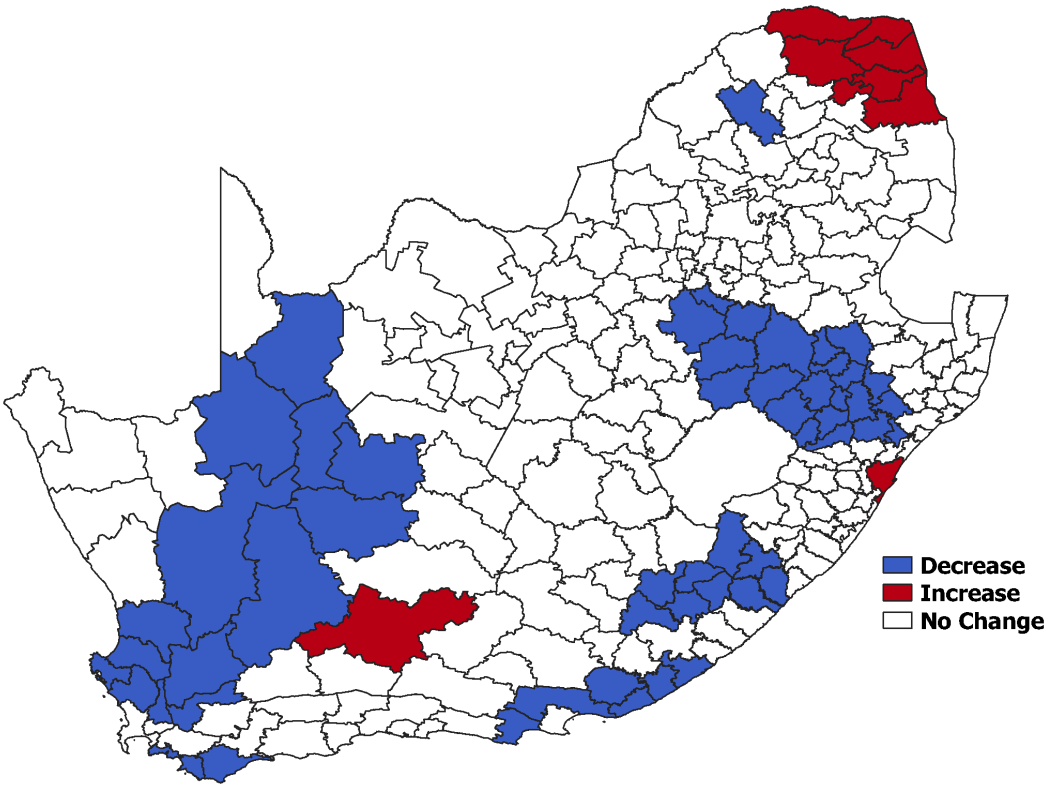


Figure 6.11: Differences between current and new characteristic wind speed maps.

6.4 Conclusion

The South African Weather Service has collected a large quantity of wind speed data which has been organised into a dataset appropriate for assessment of the South African wind loading standard. This data can be incorporated into the standard by directly considering the design values rather than characteristic values to utilise more information and better attain the target reliability.

This study does not aim to modify the reliability level of the standard. Therefore, to avoid recalibration, the annual design wind speed exceedance probability currently provided by the standard could be used to estimate design wind speeds.

To find this exceedance probability, an approach used for the American ASCE standard was investigated. This was found to be inappropriate as it made the assumption that all wind loading uncertainties could be attributed to wind speeds, which does not match the assumptions used to derive the current loads in the South African standard.

Instead, the wind load uncertainty models used to calibrate the current standard were applied to calculate the exceedance probability of a design wind speed and an effective partial factor for other components of the design load. Together these were used to estimate the design wind load using the available data and an optimal procedure that combines wind speed data collected at a specific site with regional averages.

Deviation of the estimated design values from the current design values provides evidence in favour of changing the current design values, although the strength of this evidence is dependent on the quantity and quality of the data used. To determine where there was sufficient evidence that the current design values were inappropriate, a statistical hypothesis test was developed.

This test indicated that changes to the current design values could be considered at a number of sites throughout the country. These changes were implemented in a new characteristic map of basic wind speed (Figure 6.10) that should lead to a better reflection of the South African climate in the standard.

Appendix 6

6.A Station Design Wind Speeds

Station	Latitude	Longitude	v_{50}	v_d	r	Exposure	Municipal Code
ADDO	-33.4401	25.7501	40	39.86	0.96	Accept	EC106
ALEXANDERBAAI	-28.5704	16.5282	32	34.86	0.94	Accept	NC061
ALIWAL_NORTH	-30.8029	26.8834	40	42.66	0.80	Accept	EC143
AUGRABIES_FALLS	-28.5950	20.3401	40	38.79	1.00	Accept	NC082
BABANANGO	-28.3644	31.2065	36	42.07	0.52	Accept	KZN266
BELFAST	-25.6917	30.0344	36	35.39	0.95	Accept	MP314
BETHLEHEM_WO	-28.2491	28.3337	40	39.53	1.00	Accept	FS192
BEUFORT_WEST	-32.3476	22.5733	36	46.53	0.00	Accept	WC053
BIRD_ISLAND	-33.8388	26.2861	40	34.74	1.00	Accept	EC106
BISHO	-32.8945	27.2868	40	41.11	0.99	Accept	BUF
BLOEMFONTEIN_STAD	-29.1205	26.1875	40	44.23	0.72	Reject	MAN
BLOEMFONTEIN_WO	-29.1040	26.2980	40	42.46	0.97	Accept	MAN
BLOEMHOF	-27.6511	25.6219	40	40.54	0.96	Accept	NW396
BRANDVLEI	-30.4648	20.4785	36	36.93	0.99	Accept	NC065
CALVINIA_WO	-31.4820	19.7617	36	37.91	0.95	Accept	NC065
CAPE_ST_FRANCIS	-34.2124	24.8357	40	41.58	0.99	Accept	EC108
CAPE_TOWN_WO	-33.9787	18.5998	40	40.99	0.99	Accept	CPT
CAPE_TOWN_YACHT	-33.9206	18.4430	40	56.55	0.00	Reject	CPT
CERES	-32.9635	19.4307	40	32.49	1.00	Reject	WC022
CHARTERS_CREEK	-28.1978	32.4141	32	37.58	0.28	Accept	KZN275

CLANWILLIAM	-32.1768	18.8880	36	35.59	0.95	Reject	WC012
CRADOCK	-32.1675	25.6175	40	44.40	0.81	Accept	EC131
DE_AAR_WO	-30.6651	23.9926	40	51.22	0.11	Accept	NC073
DOHNE	-32.5276	27.4606	40	38.29	0.98	Reject	EC124
DURBAN_WO	-29.9641	30.9560	36	43.94	0.05	Accept	ETH
EAST_LONDON_WO	-33.0359	27.8298	40	41.69	1.00	Accept	BUF
ELLIOT	-31.3363	27.8404	44	44.86	0.92	Accept	EC138
ELLISRAS	-23.6772	27.7069	36	36.66	0.98	Accept	GT421
ERMELO	-26.4976	29.9838	36	38.79	0.87	Accept	MP302
FORT_BEUFORT	-32.7881	26.6294	40	43.43	0.84	Accept	EC127
GARIEP_DAM	-30.5621	25.5278	40	46.89	0.47	Accept	FS162
GEELBEK	-33.1963	18.1240	40	32.19	1.00	Accept	WC014
GEORGE_WITFONTEIN	-33.9353	22.4269	36	41.20	0.71	Accept	WC044
GEORGE_WO	-34.0058	22.3893	36	39.95	0.90	Accept	WC044
GIANTS_CASTLE_AWS	-29.2700	29.5200	40	45.28	0.49	Reject	KZN236
GLEN_COLLEGE	-28.9420	26.3258	40	45.90	0.63	Accept	MAN
GRAAF-REINET	-32.1928	24.5435	36	37.91	0.98	Accept	EC101
GRAHAMSTOWN	-33.2907	26.5027	40	39.38	0.98	Accept	EC104
GRASKOP_AWS	-24.9311	30.8436	36	37.04	0.94	Accept	MP321
GREYTOWN	-29.0832	30.6033	40	39.56	0.99	Reject	KZN245
HERMANUS	-34.4323	19.2246	44	42.62	1.00	Accept	WC032
HOEDSPRUIT	-24.3544	31.0498	36	40.21	0.66	Accept	LIM335
IRENE	-25.9105	28.2106	36	42.42	0.26	Accept	TSH
IXOPO	-30.1525	30.0747	40	35.10	1.00	Reject	KZN434
JAMESTOWN	-31.1210	26.8099	40	33.56	1.00	Accept	EC143
JHB_BOT_TUIN	-26.1603	27.9981	36	33.34	1.00	Reject	JHB
JOHANNESBURG	-26.1436	28.2366	36	42.27	0.42	Accept	EKU
JOURBETINA	-33.8334	23.8644	36	38.86	0.72	Reject	EC109
KATHU	-27.6707	23.0061	36	37.92	0.95	Accept	NC453
KIMBERLY_WO	-28.8008	24.7671	40	43.54	0.84	Accept	NC091
KNYSNA	-34.0480	23.0789	36	38.47	0.88	Accept	WC048

KOINGNAAS	-30.1955	17.2901	32	30.84	1.00	Accept	NC064
KOKSTAD	-30.5028	29.3942	40	42.83	0.88	Accept	KZN433
KOMATIDRAAI	-25.5138	31.9123	32	37.88	0.34	Accept	MP324
KROONSTAD	-27.6665	27.3136	40	39.60	0.99	Reject	FS201
LADISMITH	-33.5006	21.2653	36	40.71	0.50	Reject	WC041
LADYSMITH	-28.5748	29.7505	40	43.36	0.96	Accept	KZN232
LAINGSBURG	-33.1908	20.8631	40	35.47	1.00	Reject	WC051
LAMBERTSBAAI	-32.0356	18.3313	36	29.74	1.00	Accept	WC012
LANGEBAANWEG	-32.9725	18.1577	40	38.04	1.00	Accept	WC014
LANGGEWENS	-33.2763	18.7051	40	42.47	0.84	Accept	WC015
LICHTENBURG	-26.1330	26.1644	36	39.78	0.77	Accept	NW384
LYDENBURG	-25.1119	30.4766	36	43.60	0.34	Reject	MP321
MACHADORP_AWS	-25.7158	30.2303	36	39.46	0.73	Accept	MP314
MADIKWE	-24.6884	26.1964	36	41.64	0.56	Accept	NW385
MAFIKENG	-25.8000	25.5406	36	38.00	0.91	Accept	NW383
MALMESBURY	-33.4725	18.7185	40	31.43	1.00	Accept	WC015
MANDINI	-29.1582	31.4023	36	39.93	0.76	Accept	KZN291
MARGATE	-30.8588	30.3432	36	39.73	0.86	Accept	KZN216
MARKEN	-23.5959	28.3878	36	35.88	0.95	Accept	LIM367
MBAZWANA	-27.4775	32.5983	32	38.39	0.31	Accept	KZN271
MOLTEN_RESERVOIR	-33.9375	18.4108	40	46.21	0.45	Reject	CPT
MOOI_RIVER	-29.2180	30.0025	40	39.83	0.99	Accept	KZN223
MOSSEL_BAY	-34.1894	22.1318	36	40.39	0.61	Reject	WC043
MT EDGECOMBE	-29.7068	31.0459	36	36.12	1.00	Reject	ETH
MTUNZINI	-28.9473	31.7079	36	39.59	0.92	Accept	KZN284
NEWCASTLE	-27.7688	29.9789	40	41.88	0.95	Accept	KZN252
NOUPOORT	-31.1864	24.9611	40	43.11	0.87	Accept	NC072
ORIBI_AIRPORT	-29.6479	30.4003	40	51.98	0.18	Accept	KZN225
OUDESTAD	-25.1815	29.3390	36	45.62	0.21	Accept	LIM472
PAARL	-33.7219	18.9722	40	37.94	1.00	Accept	WC023
PADDOCK	-30.7545	30.2582	36	38.58	0.94	Accept	KZN215

PATENSIE	-33.7667	24.8246	40	41.13	0.92	Accept	EC108
PENNINGTON_SOUTH	-30.3997	30.6858	36	41.47	0.61	Accept	KZN212
PIETERMARITZBERG	-29.6277	30.4034	40	39.63	0.97	Accept	KZN225
PIETERSBURG	-23.8577	29.4521	36	45.14	0.09	Accept	LIM354
PILANESBERG	-25.2538	27.2187	36	40.24	0.72	Accept	NW375
PLETTENBERGBAAI	-34.0884	23.3230	36	41.90	0.54	Accept	WC047
PONGOLA	-27.4139	31.5923	32	39.53	0.32	Accept	KZN262
PORT ELIZABETH	-33.9865	25.6168	40	48.98	0.22	Accept	NMA
PORT_ALFRED	-33.5595	26.8809	40	36.47	1.00	Accept	EC105
PORT_EDWARD	-31.0579	30.2279	36	43.17	0.41	Accept	KZN216
PORT_NOLLOTH	-29.2500	16.8683	32	25.92	1.00	Reject	NC061
PORTERVILLE	-33.0127	18.9772	40	35.79	1.00	Accept	WC013
POSTMASBURG	-28.3454	23.0790	36	40.71	0.74	Accept	NC085
POTCHEFSTROOM	-26.7359	27.0755	36	42.65	0.34	Reject	NW402
POTGIETERSRUS	-24.1936	29.0031	36	40.04	0.62	Reject	LIM367
PRETORIA_UNISA	-25.7655	28.2033	36	47.38	0.01	Reject	TSH
PRIESKA	-29.6725	22.7357	40	41.45	0.99	Accept	NC077
QUEENSTOWN	-31.9179	26.8753	44	45.09	0.98	Accept	EC134
RICHARDS_BAY	-28.7378	32.0934	36	36.75	0.95	Accept	KZN282
RIVERVIEW	-28.4440	32.1821	32	36.63	0.56	Reject	KZN275
ROBBENEILAND	-33.7989	18.3745	40	30.53	1.00	Accept	CPT
RUSTENBURG	-25.6605	27.2320	36	42.72	0.26	Accept	NW373
SOMERSET_EAST	-33.0497	25.7193	36	52.64	0.06	Accept	EC102
SPRINKBOK_WO	-29.6695	17.8793	32	43.51	0.00	Reject	NC062
STILBAAI	-34.3691	21.3911	36	39.78	0.94	Accept	WC042
STRAND	-34.1411	18.8489	40	48.98	0.21	Accept	CPT
STRUISBAAI	-34.8002	20.0569	44	41.29	1.00	Accept	WC033
TAUNG	-27.5458	24.7693	40	41.72	0.96	Accept	NW394
THABAZIMBI	-24.5781	27.4149	36	42.92	0.32	Reject	LIM361
THOYOYANDOU	-23.0800	30.3837	32	43.86	0.01	Accept	LIM344
TOSCA	-25.8780	23.9661	36	36.50	0.90	Accept	NW397

TSITSIKAMA	-34.0228	23.8985	36	38.41	0.87	Accept	EC109
TYGERHOEK	-34.1496	19.9022	40	45.62	0.64	Accept	WC031
TZANEEN-WESTFALIA	-23.7367	30.1127	36	28.03	1.00	Reject	LIM333
UITENHAGE	-33.7142	25.4351	40	44.48	0.82	Accept	NMA
ULUNDI	-28.3128	31.4206	36	38.61	0.95	Accept	KZN266
UMTHATHA	-31.5496	28.6739	44	45.97	0.96	Accept	EC157
UNI_OF_ZULULAND	-28.8517	31.8529	36	36.47	0.99	Reject	KZN282
UPINGTON	-28.4073	21.2568	40	43.39	0.96	Accept	NC083
VAN_REENEN	-28.3785	29.3861	40	39.96	1.00	Reject	KZN232
VAN-ZYLS_RUS	-26.8767	22.0472	36	30.04	1.00	Reject	NC451
VEREENIGING	-26.5687	27.9583	36	40.32	0.74	Accept	GT421
VIOOLSDRIF	-28.7694	17.6234	32	31.56	0.97	Reject	NC062
VIRGINIA	-29.7755	31.0557	36	51.85	0.02	Accept	ETH
WARMBAD	-24.8991	28.3238	36	48.97	0.04	Accept	LIM366
WELKOM	-27.9946	26.6659	40	47.08	0.32	Accept	FS184
WEPENER	-29.9156	26.8471	40	37.05	1.00	Reject	FS164
WITBANK	-25.8377	29.1918	36	41.64	0.45	Accept	EC136
WONDERBOOM	-25.6631	28.2166	36	49.81	0.07	Accept	TSH
WORCESTER	-33.6640	19.4187	44	45.45	0.99	Accept	WC025

Table 6.A1: Design wind speeds at South African weather stations. The r value is the proportion of simulated design values based on the newer data that exceed the current design value. The r value thus indicates the evidence for an increase of a decrease in the design wind speed (see Eq. 6.28). The Exposure column refers to whether the station had acceptable or unacceptable exposure for consideration in the updated map.

6.B Adjusting the characteristic value to different return periods in SANS 10160-3

The characteristic value of the basic wind speed specified in standards v_{50} is typically a 50-year return period value, i.e. a wind speed with an annual exceedance probability $p = 0.02$ (SANS 10160-3:2019; EN 1991-1-4; Holmes, 2018). The 50-year return period

corresponds to the typical design working life of a structure.

Provision is made for different design working lives by adjustment of v_{50} to a wind speed with a different return period v_k as

$$v_k = c_{prob} v_{50}. \tag{6.B1}$$

In the current South African wind loading standard SANS 10160-3:2019, c_{prob} is

$$c_{prob:sans} = \left[\frac{1 - 0.2\ln(-\ln(1-p))}{1 - 0.2\ln(-\ln(0.98))}\right]^{0.5}. \tag{6.B2}$$

The American ASCE 7-05 standard defined c_{prob} as

$$c_{prob:asce} = \left[0.36 + 0.1\ln(12/p)\right]. \tag{6.B3}$$

The expressions result in a significant discrepancy if the method described by Cook *et al.* (2011) is used to derive the implied exceedance probability of design wind speeds, this is demonstrated in Figure 6.B1.

The c_{prob} includes an assumed distribution of wind speed v. The $c_{prob:asce}$ effectively uses a Gumbel distribution with a coefficient of variation $\delta_v \approx 0.2$ to describe the distribution of v. The $c_{prob:sans}$ does not assume that v is Gumbel distributed, instead v^2 is assumed to be Gumbel distributed with a coefficient of variation $\delta_{v^2} \approx 0.23$. This is equivalent to a $\delta_v \approx 0.12$ and means that $c_{prob:asce}$ stipulates around 80% more wind speed variability than $c_{prob:sans}$. This and the difference in shape resulting from the assumption that v instead of v^2 is Gumbel distributed result in the discrepancy.

While a single distribution of v, and hence c_{prob} expression, is unlikely to be appropriate for the entire region covered by a standard, the c_{prob} should aim to be as representative as possible for the applicable region.

The $c_{prob:sans}$ was taken from EN 1991-1-4 which defines

$$c_{prob:euro} = \left[\frac{1 - K\ln(-\ln(1-p))}{1 - K\ln(-\ln(0.98))}\right]^{n}. \tag{6.B4}$$

This assumes that $v^{\overline{w}}$ is Gumbel distributed, where K can be related to the δ_v as

$$K \approx \frac{\sqrt{6}w\delta_v}{\pi - \sqrt{6}\gamma w\delta_v} \tag{6.B5}$$

and $n = 1/w$. Within this flexible format $c_{prob:asce}$ (6.B3) can be represented with $n = 1$ and $K = 0.17$.

EN 1991-1-4 states that the values for K and n should be given in the relevant National Annex and recommend that $n = 0.5$ and $K = 0.2$, based on typical European conditions, be adopted. When EN 1991-1-4 provisions were incorporated into SANS (Goliger, 2007), this recommendation was adopted and led to $K = 0.2$ and $n = 0.5$ in $c_{prob:sans}$. More research has now been conducted on the South African extreme wind climate by Kruger *et al.* (2011); Botha *et al.* (2018*b*) and in Chapter 4 and Chapter 5.

Botha *et al.* (2018*b*) and Chapter 4 recommend that $\delta_v = 0.15$ and $w = 1.6$ be considered representative of typical South African conditions. Incorporation of these recommendations would mean that $n = 0.63$ and that, by chance, the $K = 0.2$ should be maintained, resulting in an updated $c_{prob:sans}$ of

$$c^*_{prob:sans} = \left[\frac{1 - 0.2\ln(-\ln(1-p))}{1 - 0.2\ln(-\ln(0.98))}\right]^{0.63}. \tag{6.B6}$$

The effect of using $c^*_{prob:sans}$ on the implied exceedance probability of design wind speeds is shown in Figure 6.B1, and would result in values between $c_{prob:sans}$ and $c_{prob:asce}$.

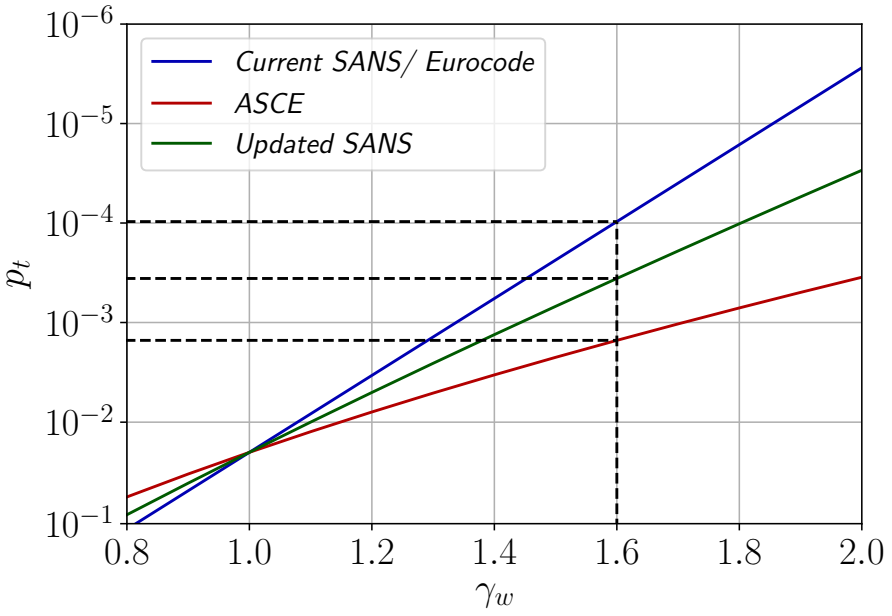


Figure 6.B1: Relationship between exceedance probability and partial factor, using the method described by Cook *et al.* (2011).

Chapter 7

Conclusion

The forces generated by wind-storms can be considered the dominant environmental load for most structures in South Africa (Kruger *et al.*, 2013*a*). The South African wind loading standard (SANS 10160-3) therefore plays a central role in structural designs throughout the country. This justifies detailed consideration of the model used by the standard, which requires an estimate of a design wind speed at a particular location.

The complexity of the wind climate means that deterministic models are of limited value, so wind speeds are estimated by using stochastic models and available data to make inferences. This study has assembled and utilised new data (Chapter 3), reviewed the spatiotemporal stochastic model (Chapter 4 and Chapter 5), and made recommendations for design loading that could be considered in the next revision of SANS 10160-3 (Chapter 6).

7.1 New Data

The current wind speed predictions in SANS 10160-3:2019 are based on research conducted by Kruger (2011), who extracted a dataset from a database managed by the South African Weather Service (SAWS). This database is continuously updated with new wind speed measurements from a nationwide network of anemometers.

In Chapter 3, Kruger (2011)'s dataset was expanded using the SAWS database to

include the subsequent 10 years (2008-2018). This study also incorporated previously unused data from 40 weather stations and measurements made between 1948 and 1990 using the now outdated Dines pressure tube anemometer. These efforts resulted in a substantial increase in the available data from around 1500 to 3500 annual 3-second gust maxima, measured at 132 stations throughout South Africa.

To standardise this data, a surface roughness correction was required. Kruger (2011) used visual inspection of satellite imagery to apply correction factors from the ISO 4354 standard. In Chapter 3, an automated correction procedure was developed and applied.

7.2 Regional Stochastic Model

Despite the significantly expanded dataset, the sample size that could be collected from an individual weather station was still a limiting factor. As a result, sophisticated extreme value approaches were not considered for an individual series and inference was based on the more robust Gumbel distribution.

A strength of the expanded dataset is the number of different stations that are available. Therefore, instead of focussing on parameter estimation of individual stations, techniques using data from multiple stations to inform parameter estimation were investigated. There were two parts to this process. The first (Chapter 4) involved regional consideration of the probability distribution and the second (Chapter 5) aimed to balance regional and site estimation of the variability.

The techniques that were developed may be useful for improving fractile estimation in situations where data from many sites, with limited sample sizes, are available. This is a common problem in wind engineering and hydrology (Holmes, 2018; Hosking and Wallis, 2005).

7.2.1 Estimating the Distribution Shape

The Gumbel distribution, with its fixed shape parameter, can be considered a special case of the Generalised Extreme Value distribution (GEV). An alternative to the GEV

distribution involves raising the wind speeds by an exponent (*preconditioning*) before fitting the Gumbel distribution. In this case, by providing additional flexibility, the exponent serves as an alternative to the shape parameter. Estimating either the shape parameter of the GEV distribution, or the preconditioning exponent, using observations from an individual series in the extended dataset, was shown to introduce significant model-variance. Compared to the less flexible Gumbel distribution.

Instead, a maximum likelihood based method which estimates a regional preconditioning exponent using multiple series in a given region was developed. The estimated regional preconditioning exponent could then be applied to all the series in the region. The maximum likelihood method was employed on the entire extended dataset from Chapter 3 and an exponent of 1.6 was found to be appropriate for South Africa. In this case the region used is the entire country. Regionally preconditioning by 1.6 was then shown to have the potential to significantly decrease the overall model-bias of design wind speed predictions in South Africa, compared to fitting the Gumbel distribution to the wind speeds (regionally preconditioning by 1) or wind pressures (regionally preconditioning by 2).

7.2.2 Estimating the Variability

Regional estimation of the variability was also considered (Hosking and Wallis, 2005). This was shown to be appropriate if the probability distributions of wind speeds throughout a region was fairly homogeneous, i.e. to differ only by a constant scaling factor. Alternatively, if a sample size was large enough, estimation of the variability based only on the individual series could be considered.

By expressing the expected error using approximate expressions, it was shown that the former *regional design* was the lower variance option and the latter *site design* was the lower bias option. It was demonstrated that the preferred option was dependent on the sample size and the degree to which the average regional coefficient of variation differed from the coefficient of variation of an individual series.

Another option, based on a typical loading standard procedure of applying a partial

safety factor to a characteristic value, was also considered. This was shown to effectively combine the site and regional design options, with the exceedance probability of the characteristic value acting as a parameter controlling the relative weighting.

An expression for an *optimal design* (based on minimum mean squared error) procedure was derived by estimating the exceedance probability of the characteristic value which minimised the mean squared error. This procedure adjusted fractile estimates between a site and regional design based on the amount of data at the site and how well that data corresponded with a regional average.

7.3 Incorporation into SANS 10160-3

The final step in this study was to use the information gained from the preceding investigations to review the current wind loading standard (SANS 10160-3:2019).

To do this, a wind speed exceedance probability was estimated from the most recent reliability calibration of the standard (Botha *et al.*, 2018*a*). Estimates of this design value, using the extended dataset and regional estimation procedures, could then be compared to what is currently provided by the standard.

Consideration of design values is justified by an observed geographical trend in the coefficient of variation, which would suggest that the current partial factor may be conservative in some regions dominated by synoptic storms, where the coefficient of variation is lower than assumed by Botha *et al.* (2018*a*). By directly considering the design wind speeds some of the unnecessary conservatism in these areas should be removed.

To guide the review, a hypothesis test was developed that aims to determine the statistical significance of the difference between an estimated design value and what is provided in the standard. This approach decreases the subjectivity involved with extending estimates from weather stations and retains some of the expert knowledge used to derive the current map.

The hypothesis test was used to develop an updated characteristic wind speed map for SANS 10160-3. By recommending changes to the characteristic wind speed map, the

current format of the standard remains largely unchanged and the information gained from the preceding investigations could be more rapidly incorporated into SANS 10160-3.

7.4 Limitations

In Chapter 3 surface roughness correction was applied deterministically, despite uncertainty in estimating the surface roughness and in applying the corrections. This uncertainty could be especially high for thunderstorm winds. A probabilistic approach where factors such as the segment surface roughness, fetch, and profile shape are treated as random variables could be more appropriate. This may be particularly beneficial for gust measurements made using a Dines pressure tube where additional uncertain parameters, such as turbulence intensity and station locations, were required for correction.

Errors in the SAWS database meant that data could not be extracted without first verifying its veracity by visually inspecting the climatic record. Visual inspection was also required to classify the climatic mechanism so that the joint distribution of the independent wind speed mechanisms could be considered, as is recommended by Gomes and Vickery (1978). The labour involved in this processes restricted the quantity of data that could be extracted.

This study has been focussed on the broad characterisation of the wind environment on a national level. Therefore, it was aimed to extract the most relevant data from as many weather stations as possible. This broad focus on many stations meant that practically only annual maxima could be extracted and asymptotic extreme value distributions considered. Thus, sample size was a significant constraint which lead to the regional estimation approach in Chapter 4 and Chapter 5.

The methods that were developed here are reliant on the assumption of the Gumbel asymptotic form. While the Gumbel distribution could be justified by the Forwards Weibull form of the parent distribution, it may be less appropriate for extremely-rare events, which have been postulated to have some atmospherical upper limit (Holmes, 2018). This is often used as justification for the Reverse Weibull asymptotic form. The

regional preconditioning approach, advocated for in Chapter 4, decreases the potential error introduced in the Gumbel assumption by transforming the wind speeds to be more compatible with the Gumbel form. Additionally, the nature of extremely-rare events means that to conclusively determine at what point the Gumbel asymptote is no longer appropriate could require many more decades of observation.

The optimal design estimation procedure derived in Chapter 5 and used in Chapter 6 assumes that the method of moments parameter estimation technique is used (Chapter 2). This was done so that the errors involved could be quantified using Kendall *et al.* (1948)'s approximations. While this is not expected to significantly affect the estimation, it is unclear how an optimal design will perform using more sophisticated site estimation procedures, like maximum likelihood or graphical generalised least squares (Chapter 2).

The hypothesis test used in Chapter 6 to update the characteristic wind map, anchors the new map to the past. This means the information that was gained throughout this study may not be fully utilised in the new recommended map of characteristic wind speeds.

7.5 Future Work

Before another broad analysis of wind speeds throughout South Africa is conducted, an effort to streamline extraction and quality control of data from the SAWS database would be beneficial. If the quality control and climate classification could be automated, the quantity of available data could be increased by orders of magnitude and more sophisticated approaches considered.

The concentration of development in metropolitan areas of South Africa means SANS 10160-3 should be utilised more in these areas. Therefore, a future study focusing exclusively on built up metropolitan areas could be considered. This could include a more refined probabilistic correction method of the measurements made using Dines anemometers. A metropolitan focussed study would be able to focus in-depth on fewer stations which would allow for a wider range of methods to be considered. Justification

for this approach will continue to increase as South Africa urbanises (Turok, 2012).

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. *Second International Symposium on Information Theory, Csaki. BNPBF Csaki Budapest: Academiai Kiado.*

Ang, A.H.-S. and Tang, W.H. (1984). Probability concepts in engineering planning and design, vol. 2: Decision, risk, and reliability. *JOHN WILEY & SONS, INC., 605 THIRD AVE., NEW YORK, NY 10158, USA, 1984, 608.*

ASCE 7-16 (2017). Minimum design loads and associated criteria for buildings and other structures. Standard, American Society of Civil Engineers.

AS/NZS 1170.2 (2011). Structural design actions-Part 2: Wind actions. Standard, Australian/New Zealand Standard (AS/NZS), Joint Technical Committee BD-006, Australia/New Zealand.

Australian Government, Bureau of Meteorology (2020). Southern hemisphere tropical cyclone data portal.
Available at: `http://www.bom.gov.au/cyclone/tropical-cyclone-knowledge-centre/history/tracks/`

Ayodele, T., Jimoh, A., Munda, J. and Agee, J. (2012). Wind distribution and capacity factor estimation for wind turbines in the coastal region of South Africa. *Energy Conversion and Management*, vol. 64, pp. 614–625.

Bakker, F.P. and Viljoen, C. (2019). An Analysis of South African Wind Gust Data in the Context of the Built Environment. In: *SEMC 2019: The Seventh International*

Conference on Structural Engineering, Mechanics and Computation, pp. 2352–2358. CRC Press.

Balakrishnan, N. and Chan, P. (1992). Order statistics from extreme value distribution, I: Tables of means, variances and covariances. *Communications in Statistics-Simulation and Computation*, vol. 21, no. 4, pp. 1199–1217.

Baravalle, M. and Köhler, J. (2018). On the probabilistic representation of the wind climate for calibration of structural design standards. *Structural safety*, vol. 70, pp. 115–127.

Beljaars, A. (1987). The influence of sampling and filtering on measured wind gusts. *Journal of Atmospheric and Oceanic Technology*, vol. 4, no. 4, pp. 613–626.

Benaroya, H., Han, S.M. and Nagurka, M. (2005). *Probability models in engineering and science*. CRC Press.

Botha, J. (2016). *Probabilistic models of design wind loads in South Africa*. Ph.D. thesis, Stellenbosch: Stellenbosch University.

Botha, J., Retief, J. and Viljoen, C. (2018*a*). Reliability Assessment of the South African Wind Load Design Formulation. *Journal of the South African Institution of Civil Engineering*, vol. 60, no. 3, pp. 30–40.

Botha, J., Retief, J.V. and Viljoen, C. (2018*b*). Uncertainties in the South African wind load design formulation. *Journal of the South african institution of civil engineering*, vol. 60, no. 3, pp. 16–29.

Brock, F.V., Richardson, S.J., Richardson, S.J. *et al.* (2001). *Meteorological measurement systems*. Oxford University Press, USA.

Burnham, K.P. and Anderson, D.R. (2002). A practical information-theoretic approach. *Model selection and multimodel inference, 2nd ed. Springer, New York*.

Castillo, E., Hadi, A.S., Balakrishnan, N. and Sarabia, J.-M. (2005). *Extreme value and related models with applications in engineering and science*. Wiley Hoboken, NJ.

Cook, N. (1982). Towards better estimation of extreme winds. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 9, no. 3, pp. 295–323.

Cook, N. (1985). The Designers Guide to Wind Loading of Building Structures Part 1. Building Research Establishment, Dept. of the Environment.

Cook, N.J. (2018). Further discussion of "The annual rate of independent events for the analysis of extreme wind speed" by Alessio Torrielli, Maria Pia Repetto & Giovanni Solari. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 174, pp. 458–463.

Cook, N.J. and Harris, R.I. (2004). Exact and general FT1 penultimate distributions of extreme wind speeds drawn from tail-equivalent Weibull parents. *Structural Safety*, vol. 26, no. 4, pp. 391–420.

Cook, N.J. and Harris, R.I. (2008). Postscript to "Exact and general FT1 penultimate distributions of extreme wind speeds drawn from tail-equivalent Weibull parents". *Structural Safety*, vol. 30, no. 1, pp. 1–10.

Cook, N.J., Harris, R.I. and Whiting, R. (2003). Extreme wind speeds in mixed climates revisited. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 91, no. 3, pp. 403–422.

Cook, R., Griffis, L., Vickery, P. and Stafford, E. (2011). ASCE 7-10 wind loads. In: *Structures Congress 2011*, pp. 1440–1453.

Cramér, H. (1946). *Mathematical methods of statistics (PMS-9)*, vol. 9. Princeton University Press.

Davenport, A. (1982). The interaction of wind and structures. *Engineering meteorology*, pp. 527–572.

Davenport, A.G. (1964). Note on the distribution of the largest value of a random function with application to gust loading. *Proceedings of the Institution of Civil Engineers*, vol. 28, no. 2, pp. 187–196.

Department of Environmental Affairs Republic of South Africa (2020). National Land-Cover Datasets.
Available at:
`https://egis.environment.gov.za/sa_national_land_cover_datasets`

Department of Rural Development and Land Reform, Republic of South Africa (2020). National aerial photography and imagery programme.
Available at: `http://www.ngi.gov.za/index.php/what-we-do/aerial-photography-and-imagery`

Department of the Environment, Heritage and Local Government (2009). Irish National Annex to the Wind Eurocode (EN 1991-1-4), Derivation of the Wind Map.

Ellingwood, B.R. and Tekie, P.B. (1999). Wind load statistics for probability-based structural design. *Journal of Structural Engineering*, vol. 125, no. 4, pp. 453–463.

EN 1990 (2002). Eurocode. Basis of Structural Design, 2002. Standard, Comité Européen de Normalisation, CEN.

EN 1991-1-4 (2005). Eurocode. Actions on structures Part 1-4: General actions-wind actions. Standard, Comité Européen de Normalisation, CEN.

Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 180–190. Cambridge University Press.

Friedman, J., Hastie, T. and Tibshirani, R. (2001). *The elements of statistical learning*, vol. 1. Springer series in statistics New York.

Fujita, T.T. (1971). Proposed characterization of tornadoes and hurricanes by area and intensity.

Goel, N., Burn, D.H., Pandey, M.D. and An, Y. (2004). Wind quantile estimation using a pooled frequency analysis approach. *Journal of wind engineering and industrial aerodynamics*, vol. 92, no. 6, pp. 509–528.

Goliger, A.M. (2007). South African wind loading specifications: The Euro way? *Journal of wind engineering and industrial aerodynamics*, vol. 95, no. 9-11, pp. 1053–1064.

Goliger, A.M. and Retief, J. (2007). Severe wind phenomena in Southern Africa and the related damage. *Journal of wind engineering and industrial aerodynamics*, vol. 95, no. 9-11, pp. 1065–1078.

Goliger, A.M., Retief, J.V. and Kruger, A.C. (2017). Review of climatic input data for wind load design in accordance with SANS 10160-3. *Journal of the South African Institution of Civil Engineering*, vol. 59, no. 4, pp. 2–11.

Goliger, A.M.W. (2016). *Wind Engineering Science and its role in optimising the design of the built environment*. Ph.D. thesis, Stellenbosch: Stellenbosch University.

Gomes, L. and Vickery, B. (1978). Extreme wind speeds in mixed wind climates. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 2, no. 4, pp. 331–344.

Gringorten, I.I. (1963). A plotting rule for extreme probability paper. *Journal of Geophysical Research*, vol. 68, no. 3, pp. 813–814.

Gulvanessian, H., Calgaro, J.-A. and Holickỳ, M. (2012). *Designers' guide to Eurocode: basis of structural design, EN 1990*. London: ICE,.

Gulvanessian, H. and Holicky, M. (2005). Eurocodes: using reliability analysis to combine action effects. *Proceedings of the Institution of Civil Engineers-Structures and Buildings*, vol. 158, no. 4, pp. 243–252.

Gumbel, E.J. (1958). Statistics of Extremes. *Press, New York*, p. 201.

Harris, R. (1996). Gumbel re-visited-a new look at extreme value statistics applied to wind speeds. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 59, no. 1, pp. 1–22.

Harris, R.I. (2017). Discussion of "The annual rate of independent events for the analysis of extreme wind speed" by Alessio Torrielli, Maria Pia Repetto & Giovanni Solari. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 100, no. 164, pp. 174–178.

Harris, R.I. and Cook, N.J. (2014). The parent wind speed distribution: Why Weibull? *Journal of wind engineering and industrial aerodynamics*, vol. 131, pp. 72–87.

Hasofer, A.M. and Lind, N.C. (1974). Exact and invariant second-moment code format. *Journal of the Engineering Mechanics division*, vol. 100, no. 1, pp. 111–121.

Holickỳ, M. (2009). *Reliability analysis for structural design*. African SUN Media.

Holmes, J. and Ginger, J. (2012). The gust wind speed duration in AS/NZS 1170.2. *Australian Journal of Structural Engineering*, vol. 13, no. 3, pp. 207–217.

Holmes, J.D. (2018). *Wind loading of structures*. CRC Press.

Hong, H. (2013). Selection of regressand for fitting the extreme value distributions using the ordinary, weighted and generalized least-squares methods. *Reliability Engineering & System Safety*, vol. 118, pp. 71–80.

Hong, H. (2015). Application of the Box-Cox Power Transformation in Extreme Value Analysis of Wind Speed. In: *12th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP12*.

Hong, H., Li, S. and Mara, T. (2013). Performance of the generalized least-squares method for the Gumbel distribution and its application to annual maximum wind speeds. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 119, pp. 121–132.

Hong, H., Mara, T., Morris, R., Li, S. and Ye, W. (2014). Basis for recommending an update of wind velocity pressures in Canadian design codes. *Canadian Journal of Civil Engineering*, vol. 41, no. 3, pp. 206–221.

Hong, H. and Ye, W. (2014). Estimating extreme wind speed based on regional frequency analysis. *Structural Safety*, vol. 47, pp. 67–77.

Hong, H.P., Ye, W. and Li, S.H. (2016). Sample size effect on the reliability and calibration of design wind load. *Structure and Infrastructure Engineering*, vol. 12, no. 6, pp. 752–764.

Hosking, J.R.M. and Wallis, J.R. (2005). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press.

Hurvich, C.M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, vol. 76, no. 2, pp. 297–307.

ISO 4354:2009 (2009). Wind actions on structures. Standard, International Organization for Standardization, Geneva, CH.

JCSS Model Code (2001 February). Probabilistic Model Code. Standard, Joint Committee on Structural Safety.

Katebi, J., Shoaei-parchin, M., Shariati, M., Trung, N.T. and Khorami, M. (2019). Developed comparative analysis of metaheuristic optimization algorithms for optimal active control of structures. *Engineering with Computers*, pp. 1–20.

Kendall, M.G. *et al.* (1948). The advanced theory of statistics. Vols. 1. *The advanced theory of statistics. Vols. 1.*, vol. 1, no. Ed. 4.

Kjeldsen, T.R., Smithers, J. and Schulze, R. (2002). Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method. *Journal of hydrology*, vol. 255, no. 1-4, pp. 194–211.

Knutson, T.R., McBride, J.L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J.P., Srivastava, A. and Sugi, M. (2010). Tropical cyclones and climate change. *Nature geoscience*, vol. 3, no. 3, pp. 157–163.

Kruger, A. (2018). Personal communication.

Kruger, A., Goliger, A.M., Retief, J. and Sekele, S. (2010). Strong wind climatic zones in South Africa.

Kruger, A., Retief, J. and Goliger, A.M. (2011). An updated description of the strong-wind climate of South Africa.

Kruger, A., Retief, J. and Goliger, A.M. (2013*a*). Strong winds in South Africa: Part 1 Application of estimation methods. *Journal of the South African Institution of Civil Engineering*, vol. 55, no. 2, pp. 29–45.

Kruger, A., Retief, J. and Goliger, A.M. (2013*b*). Strong winds in South Africa: Part 2 Mapping of updated statistics. *Journal of the South African Institution of Civil Engineering*, vol. 55, no. 2, pp. 46–58.

Kruger, A.C. (2011). *Wind climatology of South Africa relevant to the design of the built environment*. Ph.D. thesis, Stellenbosch: University of Stellenbosch.

Kruger, A.C., Retief, J. and Goliger, A. (2017). Development of an updated fundamental basic wind speed map for SANS 10160-3. *Journal of the South African Institution of Civil Engineering*, vol. 59, no. 4, pp. 12–25.

Laio, F., Di Baldassarre, G. and Montanari, A. (2009). Model selection techniques for the frequency analysis of hydrological extremes. *Water Resources Research*, vol. 45, no. 7.

Lechner, J.A., Leigh, S.D. and Simiu, E. (1992). Recent approaches to extreme value estimation with application to wind speeds. Part I: The Pickands method. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 41, no. 1-3, pp. 509–519.

Lieblein, J. (1976). Efficient methods of extreme-value methodology. Tech. Rep..

Masters, F.J., Vickery, P.J., Bacon, P. and Rappaport, E.N. (2010). Toward objective, standardized intensity estimates from surface wind speed observations. *Bulletin of the American Meteorological Society*, vol. 91, no. 12, pp. 1665–1682.

McAllister, T.P., Wang, N. and Ellingwood, B.R. (2018). Risk-Informed Mean Recurrence Intervals for Updated Wind Maps in ASCE 7-16. *Journal of structural engineering (New York, NY)*, vol. 144, no. 5.

Milford, R. (1985*a*). Development of load factors for inclusion in SABS 0160. Tech. Rep., Centre for Scientific Innovation and Research.

Milford, R. (1985*b*). Extreme value analysis of South African gust speed data. *Unpublished Internal Report*, vol. 85, no. 4.

Milford, R. (1987*a*). Annual maximum wind speeds from parent distribution functions. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 25, no. 2, pp. 163–178.

Milford, R., Goliger, A. and Vinsen, J. (1994). Tornado activity in South Africa. *Journal of the South African Institution of Civil Engineers*, vol. 36, no. 1, pp. 17–23.

Milford, R.V. (1987*b*). Annual maximum wind speeds for South Africa. *Civil Engineering/Siviele Ingenieurswese*, vol. 1987, pp. 15–19.

Miller, C., Holmes, J., Henderson, D., Ginger, J. and Morrison, M. (2013). The response of the Dines anemometer to gusts and comparisons with cup anemometers. *Journal of Atmospheric and Oceanic Technology*, vol. 30, no. 7, pp. 1320–1336.

Miller, Z. (2020). South African Tornadoes 1905 - 2019. Available at: `https://sawx.co.za/resources/history-tornadoes-south-africa/`

Modarres, R. (2008). Regional maximum wind speed frequency analysis for the arid and semi-arid regions of Iran. *Journal of Arid Environments*, vol. 72, no. 7, pp. 1329–1342.

National Oceanic and Atmospheric Administration (2020). Storm events database. Available at: `https://www.ncdc.noaa.gov/stormevents/`

Nelder, J.A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, vol. 7, no. 4, pp. 308–313.

Palutikof, J., Brabson, B., Lister, D. and Adcock, S. (1999). A review of methods to calculate extreme wind speeds. *Meteorological applications*, vol. 6, no. 2, pp. 119–132.

Parida, B., Kachroo, R. and Shrestha, D. (1998). Regional flood frequency analysis of Mahi-Sabarmati Basin (Subzone 3-a) using index flood procedure with L-moments. *Water Resources Management*, vol. 12, no. 1, pp. 1–12.

Rackwitz, R. and Flessler, B. (1978). Structural reliability under combined random load sequences. *Computers & structures*, vol. 9, no. 5, pp. 489–494.

Retief, J.V. and Dunaiski, P. (2009). *Background to SANS 10160: Basis of structural design and actions for buildings and industrial structures*. AFRICAN SUN MeDIA.

R.M. Young Company (2005). *Meteorological Instruments: Wind Monitor Model 05103*. R.M. Young Company, 2801 Aero Park Drive, Traverse City, Michigan 49686, USA. Available at: `http://www.youngusa.com/Manuals/05103-90(M).pdf`

Rózsás, Á. and Sỳkora, M. (2016). Effect of statistical uncertainties on predicted extreme wind speeds. In: *7th International Workshop on Reliable Engineering Computing, REC2016. Computing with Polymorphic Uncertain Data*, pp. 489–503.

Safaei Pirooz, A.A. and Flay, R. (2018). Response characteristics of anemometers used in New Zealand. In: *19th Australasian Wind Engineering Society (AWES) Workshop*.

SANS 10160-1:2019 (2019). South African National Standard: Basis of Structural Design and Actions for Buildings and Industrial Structures. Part 1: Basis of structural design. Standard, South African Bureau of Standards Pretoria.

SANS 10160-3:2019 (2019). South African National Standard: Basis of Structural Design and Actions for Buildings and Industrial Structures. Part 3: Wind Actions. Standard, South African Bureau of Standards Pretoria.

Scholz, F.W. and Stephens, M.A. (1987). K-sample Anderson–Darling tests. *Journal of the American Statistical Association*, vol. 82, no. 399, pp. 918–924.

Simiu, E., Heckert, N., Filliben, J.J. and Johnson, S.K. (2001). Extreme wind load estimates based on the Gumbel distribution of dynamic pressures: an assessment. *Structural Safety*, vol. 23, no. 3, pp. 221–229.

Snyman, J.A. (2005). *Practical mathematical optimization*. Springer.

South African Weather Bureau (1950–1980). Report on Meteorological Data. Tech. Rep., Department of Transportation, Weather Bureau, Pretoria, South Africa.

South African Weather Service (2017). Verification of wind gusts for quality control purposes. Unpublished.

South African Weather Service (2020). A strong cold front will make landfall... [Online; accessed 27 September 2020].
Available at: `https://www.facebook.com/WeatherServic/photos/a.542579239278737/1339258746277445`

Torrielli, A., Repetto, M.P. and Solari, G. (2013). Extreme wind speeds from long-term synthetic records. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 115, pp. 22–38.

Torrielli, A., Repetto, M.P. and Solari, G. (2016). The annual rate of independent events for the analysis of the extreme wind speed. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 156, pp. 104–114.

Torrielli, A., Repetto, M.P. and Solari, G. (2017). Response to the Discussion on "The annual rate of independent events for the analysis of extreme wind speed, by R. Ian

Harris". *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 100, no. 164, pp. 179–181.

Turok, I. (2012). *Urbanisation and development in South Africa: Economic imperatives, spatial distortions and strategic responses*. Human Settlements Group, International Institute for Environment and Development United Nations Population Fund.

Von Mises, R. (1936). La distribution de la plus grande de n valuers. *Rev. math. Union interbalcanique*, vol. 1, pp. 141–160.

Walshaw, D. (1994). Getting the most from your extreme wind data: a step by step guide. *Journal of Research - National Institute of Standards and Technology*, vol. 99, pp. 399–399.

Wieringa, J. (1992). Updating the Davenport roughness classification. *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 41, no. 1-3, pp. 357–368.

World Meteorological Organization (2008). WMO guide to meteorological instruments and methods of observation.

Ye, W., Hong, H. and Wang, J. (2015). Comparison of spatial interpolation methods for extreme wind speeds over Canada. *Journal of Computing in Civil Engineering*, vol. 29, no. 6, p. 04014095.

Chapter Declarations

Declaration by the candidate (Chapter 3)

With regard to Chapter 3, the nature and scope of my contribution were as follows:

Nature of Contribution	Extent of Contribution
methodology, software, investigation, data collection and curation, writing, visualisation	80%

The following co-authors have contributed to Chapter 3

Name	E-mail address	Nature of Contribution	Extent of Contribution
C Viljoen	celesteviljoen@sun.ac.za	conceptualisation, data acquisition, writing - review & editing	15%
N de Koker	ndekoker@sun.ac.za	writing - review & editing	5%

Signature : ..... Date: .....
FP Bakker

Declaration by the co-authors (Chapter 3)

The undersigned hereby confirm that

1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to Chapter 3,
2. no other authors contributed to Chapter 3 besides those specified above, and
3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 3 of this dissertation.

Signature : ..... Date: .....
N de Koker

Signature : ..... Date: .....
C Viljoen

*Declaration with signatures in possession of candidate and supervisors.

Declaration by the candidate (Chapter 4)

With regard to Chapter 4, the nature and scope of my contribution were as follows:

Nature of Contribution	Extent of Contribution
conceptualisation, methodology, software, investigation, data curation, writing, visualisation	80%

The following co-authors have contributed to Chapter 4

Name	E-mail address	Nature of Contribution	Extent of Contribution
N de Koker	ndekoker@sun.ac.za	methodology, formal analysis, writing - review & editing	15%
C Viljoen	celesteviljoen@sun.ac.za	conceptualisation, writing - review & editing	5%

Signature : ..... Date: .....
FP Bakker

Declaration by the co-authors (Chapter 4)

The undersigned hereby confirm that

1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to Chapter 4,
2. no other authors contributed to Chapter 4 besides those specified above, and
3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 4 of this dissertation.

Signature : ..... Date: .....
N de Koker

Signature : ..... Date: .....
C Viljoen

*Declaration with signatures in possession of candidate and supervisors.

Declaration by the candidate (Chapter 5)

With regard to Chapter 5, the nature and scope of my contribution were as follows:

Nature of Contribution	Extent of Contribution
conceptualisation, methodology, software, investigation, data curation, writing, visualisation	85%

The following co-authors have contributed to Chapter 5

Name	E-mail address	Nature of Contribution	Extent of Contribution
N de Koker	ndekoker@sun.ac.za	methodology , writing - review & editing	10%
C Viljoen	celesteviljoen@sun.ac.za	writing - review & editing	5%

Signature : ..... Date: .....
FP Bakker

Declaration by the co-authors (Chapter 5)

The undersigned hereby confirm that

1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to Chapter 5,
2. no other authors contributed to Chapter 5 besides those specified above, and
3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 5 of this dissertation.

Signature : ..... Date: .....
N de Koker

Signature : ..... Date: .....
C Viljoen

*Declaration with signatures in possession of candidate and supervisors.

Declaration by the candidate (Chapter 6)

With regard to Chapter 6, the nature and scope of my contribution were as follows:

Nature of Contribution	Extent of Contribution
conceptualisation, methodology, software, formal analysis, investigation, data curation, writing, visualisation	80%

The following co-authors have contributed to Chapter 6

Name	E-mail address	Nature of Contribution	Extent of Contribution
N de Koker	ndekoker@sun.ac.za	methodology , writing - review & editing	10%
C Viljoen	celesteviljoen@sun.ac.za	conceptualisation, methodology, writing - review & editing	10%

Signature : ..... Date: .....
FP Bakker

Declaration by the co-authors (Chapter 6)

The undersigned hereby confirm that

1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to Chapter 6,
2. no other authors contributed to Chapter 6 besides those specified above, and
3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 6 of this dissertation.

Signature : ..... Date: .....
N de Koker

Signature : ..... Date: .....
C Viljoen

*Declaration with signatures in possession of candidate and supervisors.